REPORT RESUMES

ED 003 318 24
A COMPARISON OF THREE METHODS OF TEACHING ELEMENTARY SCHOOL SCIENCE INVOLVING PROGRAMMED LEARNING.
BY- HEDGES, WILLIAM D. MACDOUGALL, MARY A.
VIRGINIA UNIV., CHARLOTTSVILLE
REPORT NUMBER CRP-1972 PUB DATE 65
EDRS PRICE MF-$0.27 HC-$7.12 178P.

DESCRIPTORS- *SCIENCE INSTRUCTION, *PROGRAMED MATERIALS, ELEMENTARY EDUCATION, PERFORMANCE, GRADE 4, *PROGRAMED INSTRUCTION, COMPARATIVE ANALYSIS, CHARLOTTESVILLE, VIRGINIA

THIS STUDY--(1) EVALUATED THREE MAJOR METHODS OF SCIENCE INSTRUCTION, (2) ESTABLISHED ANALYSIS CRITERIA FOR CURRICULUM REVISION AT THE FOURTH-GRADE LEVEL, AND (3) INVESTIGATED THE INFLUENCE OF THREE RESPONSE FACTORS IN PROGRAMED LEARNING. A TREATMENT BY LEVELS OF ANALYSIS OF VARIANCE WAS USED TO TEST DIFFERENCES IN ACHIEVEMENT BEFORE AND AFTER THE RETENTION PERIOD. DIFFERENCES IN INTEREST AMONG THE THREE GROUPS WERE DETERMINED BY A CHI-SQUARE TEST. OTHER APPROPRIATE STATISTICS WERE USED AS MEASURES OF (1) THE EFFECT OF THE LABORATORY EXPERIENCES, (2) RESPONSE FACTORS AND INTERACTION EFFECTS, AND (3) CURRICULUM REVISION. INITIAL EXPERIMENTATION AND FIELD TESTING BY THE INVESTIGATORS SUGGESTED THAT PROGRAMED SCIENCE MATERIALS, COUPLED WITH THE INDIVIDUAL PERFORMANCE BY STUDENTS OF LABORATORY EXPERIENCES, HAD PROMISE AS ONE INSTRUCTIONAL METHOD IN THE MODERN ELEMENTARY SCHOOL CURRICULUM. (LP)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

A COMPARISON OF THREE METHODS OF TEACHING ELEMENTARY SCHOOL SCIENCE INVOLVING PROGRAMMED LEARNING

Cooperative Research Project No. 1972

William D. Hedges
Mary Ann MacDougall

University of Virginia
Charlottesville, Virginia

1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ED 003318

ACKNOWLEDGEMENTS

Reported here are the results of almost four years of investigation into the efficacy of programing techniques applied to science at the elementary school level. Unique aspects of the research include: (1) the evaluation of three major methods of science instruction, (2) the influence of three response factors in programed learning, (3) the provision throughout for each child to perform numerous experiments by himself upon the completion of certain framed sections, and (4) the introduction of elements of the critical thinking process into the programed materials.

Numbers of children and teachers were involved in the field testing. Thus, from the spring of 1962 to the summer of 1965, 51 fourth-grade classes totaling approximately 1600 students in 11 elementary schools located in two different school systems in Virginia participated for periods of time ranging from one month to one semester to a full academic year.

We are grateful to all teachers who participated. Each gave extra time and effort to the project despite their many other duties. Their names are:

Mrs. S. Almond
Mrs. S. Barksdale
Mrs. B. Boxley
Mrs. K. Bradley
Mrs. E. Burnet
Mrs. R. Byers
Mrs. M. Carpenter
Miss M. Chisholm
Miss J. Coffey
Mrs. N. Dunnavant
Mrs. B. Flora
Miss H. Hamilton
Mrs. S. Hatcher
Mrs. E. Heltzel
Mrs. J. Holloway
Mrs. L. Jarrett
Mrs. L. Johnson
Mrs. V. McIntire
Mrs. M. McKeith
Mrs. B. Meadows
Mrs. S. Morgan
Mrs. L. Phipps
Mrs. S. Riddick
Mrs. R. Shumate
Mrs. O. Smith
Mrs. F. Stafford
Mrs. S. Steinbaum
Mrs. A. Utz
Mrs. L. Veliky
Mrs. J. Vest
Mrs. J. Walker
Mrs. G. Watson
Miss G. White
Mrs. J. Williamson
Mrs. H. Willoughby
Miss E. Wood
Mrs. E. Woodford
Mrs. R. Yoder

It is important to note that the superintendents and school boards of Orange County and Charlottesville, Virginia, school systems made it possible for the research to be conducted by giving their approval. Their support is sincerely appreciated.

During the three and one-half year period of the study, a a total of 21 graduate students contributed time and effort ranging from direct participation in the classrooms in a close working relationship with the teachers to materials preparation to data analysis. Their names are:

Larry Bostian
Robert Brown
Jesse Byler
Panayiotes Constantinides
J.F. Crabtree, II
Sherman Dutton
Delores Eisenson
Catherine Eubanks
Keith Hanson
Ian Lochhead
Hugh MacDougall
Jeanne Mann
William Rodgers
Robert Rose
Robert Shaw
Paul Slayton
Ann Smothers
John Stubbings
Alton Taylor
Robert Todd
Arthur Welch

Special thanks goes to Dr. Panayiotes D. Constantinides for his devotion throughout all phases of the project.

Mrs. Judith Hockey, project secretary, also deserves particular commendation for she followed the project from its inception to the writing of the final report. Without her careful coordination of all phases of the project, ranging from materials preparation to careful filing of data, chaos would have resulted.

We are deeply indebted to the staff of the Cooperative Research Branch of the U.S. Office of Education for their encouragement and support.

William D. Hedges

Mary Ann MacDougall

Charlottesville, Virginia
July 1, 1965

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

LIST OF TABLES vi

INTRODUCTION ix

Chapter

I BACKGROUND RESEARCH LEADING TO THE STUDY. . 1

Phase I 4

Objectives. 5

Procedure 6

Findings. 8

Conclusions 10

Phase II 12

Objectives. 12

Procedure 13

Findings. 15

Conclusions 29

II PHASE III - THE EVALUATION OF THREE METHODS OF SCIENCE INSTRUCTION AT THE FOURTH GRADE LEVEL 33

Objectives. 33

Procedure 35

Findings. 37

Conclusions 52

Chapter Page

III PHASE IV - THE INFLUENCE OF THREE RESPONSE FACTORS IN PROGRAMED LEARNING 56

Objectives. 57

Procedure 58

Findings. 59

Conclusions 68

Subjective Appraisal of the Project . . . 69

Observations by Teachers. 69

Observations by Students. 71

Observations by Research Assistants . . 74

IV CURRICULUM ANALYSIS 86

V PROGRAMED MATERIALS WITH ELEMENTS OF CRITICAL THINKING 94

Statement of the Problem. 94

Method and Procedure. 98

Analysis of Data. 101

Findings and Conclusions. 110

VI CONCLUSIONS AND IMPLICATIONS. 112

Summary 119

BIBLIOGRAPHY 121

APPENDICES 124

LIST OF TABLES

Table		Page
I	Means of I.Q., Intitial and Final Achievement Test Scores and Mean Gains for Individual Classes.	9
II	Initial and Final Science Mean Achievement of the Experimental, Control, and Accelerated Groups.	16
III	Initial and Final Mean Achievement Converted to Percentile Ranks: Norms Group Based on Fall Testing, Grades 4 and 5	17
IV	Analysis of Variance of X and Y Scores Taken Separately.	19
V	Analysis of Co-Variance Summary Table . . .	20
VI	Distribution of the Responses, Dislike-Like of the Students in the Experimental and Control Groups	23
VII	Distribution of Dislike-Like Response Versus Reasons for Liking and Disliking Given by the Experimental Groups and Accelerated Class	26
VIII	Distribution of Dislike-Like Responses Versus Reasons for Liking and Disliking Given by the Control Group.	27
IX	Analysis of Variance Summary Table for Intelligence.	38
X	Analysis of Variance Summary Table for STEP Form 4A.	39
XI	Two Criteria Analysis of Variance Summary Table for Treatments by Levels (I.Q.) on STEP 4B Science Achievement in May.	41
XII	One Criterion Analysis of Variance Summary Table for Division True-False Post-Test in May.	42

Table		Page
XIII	Two Criteria Analysis of Variance Summary Table for Treatments by Levels (I.Q.) on Type of Interest in Science.	43
XIV	Two Criteria Analysis of Variance Summary Table for Treatments by Levels (I.Q.) on Level of Interest in Science.	44
XV	Means by Treatments and Levels on Laboratory Resourcefulness Test	50
XVI	Three Criteria Analysis of Variance Summary Table for Laboratory Resourcefulness Test.	50
XVII	One Criterion Analysis of Variance Summary Table for STEP 4B Retention - Following Summer Vacation	51
XVIII	One Criterion Analysis of Variance Summary Table for Division True-False Test in Retention Following Summer Vacation	52
XIX	Analysis of Variance Summary Table for Intelligence for Eight Treatments . .	60
XX	Analysis of Variance Summary Table for STEP for Eight Treatments	61
XXI	Analysis of Variance Summary Table for True-False Test for Eight Treatments.	62
XXII	Three Criteria Analysis of Variance Summary Table for Final Achievement Test (STEP)	64
XXIII	Three Criteria Analysis of Variance Summary Table for Final Achievement Test (True-False)	65
XXIV	Three Criteria Analysis of Variance Summary Table for Final Level of Interest.	66

Table		Page
XXV	Three Criteria Analysis of Variance Summary Table for Final Type of Interest.	67
XXVI	Analysis of Selected Frames on the Sound and Light Units	89
XXVII	Content Analysis of Items Most Frequently Recalled Correctly and Incorrectly on Sound Unit Test.	93
XXVIII	Significance of Difference Between the Means of the Experimental and Control Groups on Four Initial Measures	102
XXIX	A Mean Comparison Between the Experimental and Control Groups on the Final Measures of Critical Thinking Ability	103
XXX	Mean Comparison Between Pre- and Post-Experimental Means of Critical Thinking Ability for the Total Sample, the Experimental Group, and the Control Group	104
XXXI	Experimental and Control Groups Each Divided into Subsets of Higher and Lower Ability Levels with Related I.Q. and Critical Thinking Data.	105
XXXII	The Interrelationships Among the Critical Thinking, Intelligence, Personality and Science Achievement Variables, and the Means and Standard Deviations of Each Variable.	108

INTRODUCTION

The influence of scientific and technological demands of contemporary society has emphasized science *per se* and the re-evaluation of the science curriculum in the educational system. The impact on the elementary schools to provide more substantially for individual differences in rate of learning and achievement of students, has created a need for improved science instruction.

If programed learning were feasible within the context of modern elementary curriculum theory and if the nature and form of this approach to teaching areas within the curriculum were determined, this method of instruction could provide an essential adjunct to classroom learning and contribute to the development and sequencing of the elementary science curriculum. Thus, two major questions were considered: (1) is programed learning within the context of modern elementary curriculum theory feasible, and (2) if the former, which approaches might be more effective?

I. Major Objectives

Outlined below are the studies on programed learning coupled with the student performance of laboratory experiences which have been conducted by the Bureau of Educational Research, University of Virginia, during the years 1962-65.

After a pilot study (April, 1962) and field testing of the materials over a one-year period, experiments were designed and carried out in the fall of 1963 and in the spring of 1964 to evaluate the instructional effectiveness of the programed materials.* The control of content and method of presentation of the materials had immediate application to an analysis of the curriculum and to a study of certain techniques of critical thinking.

In order to assist the reader, each study is presented individually, and can be read as a single piece of research or as a part of the entire project. The order of presentation and major objectives of each research study is summarized below.

Background Research

Chapter I, "Background Research Leading to the Study," describes research efforts which preceded the research carried out under the contract with the Cooperative Research Branch, Department of Health, Education and Welfare.

Phase I. The general purpose of Phase I was to ascertain whether programed instruction at the fourth-grade level appeared to have promise as one method of teaching elementary school science by observing (a) children conducting their own experiments and measuring their achievement and rate of progress, (b) the general attitudes of students and teachers, and (c) the administrative feasibility of such an approach.

---

*Under a grant from the Cooperative Research Branch, H.E.W.

Phase II. The primary purpose of Phase II was to write and refine through field testing and analysis a complete course of study for the fourth grade, using Skinner's small step approach, his theory of cuing, a linear structure, constructed responses, and immediate reinforcement.

Evaluation

Chapter II, "The Evaluation of Three Methods of Science Instruction at the Fourth-Grade Level," describes the first major attempt to evaluate the programed materials in a controlled situation.

Phase III. The general purpose of Phase III was to compare the instructional effectiveness of teaching science at the fourth grade level using three methods: (a) programed science materials with student performance of laboratory experiences, (b) programed science materials with students reading about laboratory activities, and (c) a textbook format with teacher demonstrations of laboratory activities. Instructional effectiveness was defined as science achievement, interest and attitudes, laboratory resourcefulness, and retention of science material.

Chapter III, "The Influence of Three Response Factors in Programed Learning," presents an evaluation of the instructional effectiveness of the programed materials which have been modified to ascertain the influence of three response variables: response form, frequency of feedback, and response effect.

Phase IV. Programed texts were prepared to investigate the influence of the variables: (a) constructed response vs. multiple choice, (b) all answers vs. partial feedback, and (c) branching vs. linear programing. The versions were compared on the measures of instructional effectiveness used in Phase III of the evaluation.

Curriculum Analysis

The two major objectives of the curriculum analysis were: (a) to determine the level of difficulty for this general age group established from an analysis of student achievement, and (b) to determine which science concepts are suitable in level of difficulty, but require adjustments in methods of presentation to satisfy variations among ability levels for this age group. Chapter IV presents the curriculum information.

Critical Thinking

Chapter V is devoted to an investigation of elements of critical thinking introduced into the programed materials and the relationship of critical thinking with intelligence, achievement, and personality variables.

Summary and Implications

A summary of the major findings and conclusions of each of the studies described above is presented in Chapter VI.

II. Related Research

"Elementary school science has made an important place for itself in the school curriculum, but much remains to be done in building more effective programs." (6) From the pioneer study of objectives by Craig in 1927 (8), to such studies as Gilbert's analysis of 30 courses of study in 1943 (13), to Dubin's examination of 163 courses of study in 1953 (9), the conclusion is drawn that there is only limited agreement concerning not only what to include in the elementary science curriculum and on what grade level, but on how to teach it. More recently, Bryant (1) analyzed twenty curriculum guides suggested by a committee of the ASCD. His analysis also indicated there is little agreement in either form or content, scope or sequence.

The mode of presentation of programed materials seemed to offer the possibility of analyzing the curriculum development in providing a course of study for classroom use. Also, the paucity of research at the elementary school level in the field of science related to the feasibility of programed learning and laboratory experiences suggested that research in this area was needed. Experiments by Keislar and McNeil(17) in which first-grade children were taught scientific explanations by means of a 432-frame program, and Carpenter (2), in which the performance of fourth-grade students taught science under a more traditional approach or by the use of demonstrations and classroom experimentation, encouraged further research to evaluate the use of programed materials and laboratory experiences at the elementary level.

Numerous studies reported on the general effect or evaluation of programed materials. Among those reported, Hughes (15), Porter (19), and Pressey (23), indicate superior performance with the use of programed materials. Many studies, however, have indicated nearly equal achievement between programed and traditional groups, while the "halo" effect was frequently operating in favor of the programed learning group. Carr (3) concluded in 1962 that we are lacking in experiments "which unequivocally demonstrate the superiority of automated instruction over the 'usual' classroom procedures. And too, one might say that we know even less about _why_ automated instruction might prove to be superior."

Variation in programing methods and the investigation of programing variables were two common attempts to determine why a programing technique might be superior. Since much of the theory of programing strongly adheres to the psychological principles that one learns by active participation and that learning is facilitated by immediate reinforcement, it was decided that response factors would be investigated after the initial evaluation of the materials. The variables selected were, (a) the nature of the response, (b) the frequency of feedback to the learner, and (c) the response effect.

The nature of the response, defined in this study as multiple choice vs. constructed response, both require the learner's active participation, although Skinner (26) favors the constructed response, and Pressey (22), multiple choice items. Fry (12) recommends further investigation of the response modes, multiple choice and constructed response.

Shaping and maintaining behavior by immediate reinforcement and arrangement of contingencies of reinforcement are discussed as important factors in the field of learning (25); a study of partial vs. full feedback compares student performance under the two conditions.

The third variable, the effect of the learner's response, compares the influence of linear programing with branching. If linear programing can provide for differentials in learning rate between students, can branching then make provisions for (a) variations in initial knowledge levels and (b) variations in the "paths" which different students' minds need to select in most efficiently reaching similar goals. Although Skinner (26) recommends that the student pass through a sequence of st~~o~~eps in carefully prescribed order, several programing methods include branches to further individualize the instructional process. Initial findings by Coulson and Silberman (6) found no difference between the branching and no branching versions. Subsequent experiments (5) indicate the superiority of branching versions.

The terms "discovery," "problem-solving," "reflective thinking," or "critical thinking" describe similar areas of concern about the process of acquiring information. Suchman's findings from his training program in scientific inquiry (27) and the extensive work of Inhelder and Piaget (16) give evidence of the nature and growth of thinking ability and strongly recommend continued research into this problem. An amendment to the contract allowed the inclusion of elements of critical thinking into four units of the programed materials.

CHAPTER I
BACKGROUND RESEARCH LEADING TO THE STUDY

This project was conceived in the fall of 1961. The reader may recall that at that time "programing" and "teaching machines," were words flashing across the educational world coupled with the names of men such as Skinner, Pressey, and Crowder. Reactions of educators ranged from complete skepticism to prognostications of, for the first time, being able to provide in a realistic manner for individual variations in rate and route to the mastery of certain educational goals. The publishing world was quite interested. Unfortunately, at least one very vital element was missing, an adequate research basis with human beings. The bulk of the related research concerned animals. Research was sorely needed to answer such questions as, (1) Does Skinner's method really work with human beings? (2) If so, what is meant by 'work?' (3) How about delayed retention? (4) If effective with human beings, are there some subjects or skill areas or age levels for which it is better suited than others? (5) How about Crowder's branching concepts? (6) Are these methods better suited for college or graduate school than high school or elementary school? (7) How about the student characteristics such as intelligence; i.e., would the new methods work well with some and not with others and, if so, why?

The authors of this project had questions such as the above in mind, but they were thinking primarily of the implications, if any, of these developments for children studying elementary school science, but with one major modification: the children would perform experiments individually as they completed certain framed sections in the materials.

The field of elementary school science was in difficult straits. The intensive efforts of the National Science Foundation and other agencies bringing about extensive retraining of secondary school science teachers; the many new science curriculums such as PSSC, BSCS, etc., all had combined to bring about greater pressures since sputniks on the elementary school to upgrade the teaching of science. It came to be recognized that (1) in many schools very little science was being taught, (2) in most schools teachers felt unprepared to deal adequately with science--and said so. A study of the status of science in the elementary schools of Virginia about this time supports these statements.[1]

---

[1] William D. Hedges and Mary Ann MacDougall, "An Investigation of the Status of Science Education in Selected Public Elementary Schools of Virginia," Division of Educational Research, University of Virginia, Charlottesville, Virginia, September, 1962.

In short, the authors asked themselves questions such as the following:

1. Does programed science have a part to play in the modern elementary school science curriculum? Specifically:
    - A. Can programed science contribute toward the individualization of science instruction? If so, how?
    - B. Can programed science contribute more than new information (facts), or must its major contribution be recognized as just that--information imparting?
    - C. Is there a "best" method in general; i.e., does Skinner's approach seem to achieve better than Crowder's? Does the efficacy of either vary with student ages - student characteristics?
    - D. How about level of concept development? Can programing tell us anything about the maturation of children and optimum time for teaching certain kinds of children?
    - E. Is the individualization in rate of progress for children realistic in the regular classroom setting; i.e., can teachers handle it?
    - F. What about the interest of children? Is it high? Is it sustained?
    - G. How about the individual performance by children of their own experiments? Is this reflected in additional achievement? In motor skill development? In growth in their ability to solve problems?
    - H. Field testing of programs was for short periods. How about the efficacy of programing over a period of one semester? of an academic year?

Questions such as the above and dozens of others occurred to the investigators. As a result, they decided to investigate the efficacy of teaching fourth grade science by means of programed science materials coupled with laboratory experiences individually performed by each of the students as he completed certain framed sections. The study, as conceived, was to be broken into four phases and stretched over a period of three years. Whether or not the second, third, and fourth phases would be attempted depended on the results of the first phase.

PHASE I

The Orange County Pilot Study[2]

During the fall and early spring of 1961-62, three science units were programed using Skinner's small step method and his theory of cuing with a linear structure employing constructed responses with immediate knowledge of results. In April of 1962 these materials were field tested in two rural elementary schools of Orange County, Virginia using two fourth grade control and two experimental classes. The field testing was designed to extend over a period of approximately one month depending, of course, on how rapidly the students actually progressed through the materials.

---

[2] Sherman S. Dutton, "An Experimental Study in the Programing of Science Instruction for the Fourth Grade," (unpublished Doctoral Dissertation, University of Virginia, Charlottesville, Virginia, 1963).

A. Objectives

The general purpose of Phase I was to ascertain whether programed instruction at the fourth grade level appeared to have promise as one method of teaching elementary school science by observing (1) children conducting their own experiments, and measuring their achievement and rate of progress; (2) the general attitudes of students and teachers, as well as administrative problems; (3) the nature and extent of curriculum revisions. The specific objectives of the study were to determine:

1. Whether the mean science achievement of students using programed materials differed from the mean science achievement of students using a more conventional approach.
2. The extent to which students were actually able to and did proceed at varying rates.
3. The feasibility of allowing each fourth grade student to perform his own experiments as he arrived at the end of specific framed sections.
4. Whether students liked the materials and whether such attitude was sustained and whether it varied with students.
5. Administrative problems associated with the use of such science equipment and materials.
6. An expression of teacher reaction.
7. The nature and extent of revisions plus guidelines for writing additional materials should the pilot study indicate such seemed desirable.

B. Procedure

Materials in heat, light, and sound were programed using Skinner's small step approach and his theory of cuing with a linear structure employing constructed responses with immediate knowledge of results. These materials were administered to three fourth grade students and revisions were made including vocabulary simplification, rephrasing of statements, and addition of intermediate frames for finer increments.

Two fourth grade classes in each of two rural elementary schools were utilized, for a total of one hundred eleven students. One grade from each school was randomly selected for the experimental group of two classes (N = 53); the other two classes (N = 58) constituted the control group.

Since a standardized test peculiar to the measurement of achievement in heat, light, and sound, of the fourth grade level was not available, one had to be constructed (true-false, 101 items). This test was pre-administered to a fourth-grade class of 32 students who had studied heat, light, and sound earlier in the year. The reliability coefficient, using the split-halves method with the Spearman-Brown prophecy formula was .84. Because of the possible test bias towards the programed group, the use of the instrument was restricted to the pilot study. In the subsequent study, extended over a one-year duration, standardized tests, such as the STEP would be available.

The above-mentioned test was then administered to each of the four fourth-grade science classes at the beginning and end of the experimental period. Lorge-Thorndike Intelligence Test scores were available from students' records.

The class schedules were arranged so that each class studied science for thirty minutes each on Monday, Wednesday and Friday for a period of five weeks, a total of seven and one half hours of science instruction. It should be noted, however, that only the two control classes who were taught with fixed assignments by teachers actually studied the full seven and one half hours. Time rates in the experimental class varied with a median time of five hours. Thirteen students, using programed materials, finished in four hours or less, while three students required seven to seven and one half hours.

A one criterion analysis of variance design was used, with the criterion measure defined as final science achievement test scores. The use of an analysis of covariance, holding science achievement and I.Q. constant, was unnecessary, since differences in mean I.Q. and initial science achievement between the experimental and control groups were nonsignificant (groups were practically identical). An F-ratio reaching the .05 level was assumed to reject the null hypothesis of no mean difference in final science achievement between the experimental and control groups.

Students who finished the materials early were given work in another programed unit (electricity) to avoid a waste in their time. Each student was administered the post-test, however, immediately upon completing the programed units in heat, light and sound and before going ahead with the unit on electricity.

C. Findings

The initial mean I.Q., initial and final mean science achievement, and mean gains are presented in TABLE I. This experiment revealed the following findings:

1. The achievement of those students using programed materials for this experiment was significantly higher at the .05 level than that of those taught by certain more conventional methods.
2. The teachers who cooperated in this experiment indicated that the use of programed science materials was beneficial to them in their elementary school instruction and that they as teachers would welcome such instructional assistance.

TABLE I

MEANS OF I.Q., INITIAL AND FINAL ACHIEVEMENT TEST SCORES AND MEAN GAINS FOR INDIVIDUAL CLASSES[3]

	Control			
	Means of			
	I.Q.	Pre	Post	Gain
School A	96	57.58	74.42	16.84
School B	101.7	55.34	65.97	10.63
	Experimental			
	Means of			
	I.Q.	Pre	Post	Gain
School A	95.3	55.88	79.36	23.48
School B	101.6	58.18	76.96	18.78

---

[3] Sherman S. Dutton, "An Experimental Study in the Programing of Science Instruction for the Fourth Grade," unpublished doctoral dissertation, University of Virginia, Charlottesville, Virginia, 1962, p. 47.

3. There existed an uncontrolled variable in this experiment. The interest and enthusiasm of the classroom teacher could have affected student achievement since the different classroom groups showed considerable variation in achievement.

4. There appeared to be considerable student interest exhibited toward the use of programed materials.

5. Individual student experiments appeared to serve as an additional motivational factor for fourth grade students.

6. Fourth grade students followed written instructions and performed simple experiments with relatively little assistance from the classroom teacher. Most difficulty was encountered with the poor readers.

7. Although the fourth grade students proceeded at highly varying rates in the classroom situation, this did not present any administrative problems that could not be handled by the teacher although she was a very busy person.

8. Individual rates of progress did not separate students sufficiently and students tended to cluster on the first experiment causing some loss of time for those students who had to wait for their turn to use the equipment. This problem vanished after the first week.

D. Conclusions

The study served its major function; i.e., indicating that if an excellent course of study were to be programed, this approach to teaching science might have considerable promise because:

1. After the first week, only one set of equipment for each experiment became necessary for students no longer clustered.

2. Students were extremely interested in performing their own experiments.

3. Significant mean differences in final science achievement were in favor of the experimental group.

4. Students could and did proceed at individual rates of speed.

5. Students indicated they enjoyed the materials.

6. Teachers enjoyed the materials. As example, one teacher chose to go ahead with other units when the Pilot Research Study was completed.

Despite the statistically significant results in achievement reported above, no conclusions were drawn because of the following weaknesses in the pilot study.

1. It was of extremely short duration - only five weeks.

2. The control groups were not really control groups for what is a "conventional method of studying science." The teachers in the control groups were given the objectives of the course, a rough outline and some help with materials and equipment but comparable text book materials were not readily available. The programed materials were more comprehensive than the texts used in the control groups.

3. A halo effect of tremendous interest was introduced to the students in the experimental classes, who became more excited over a "new method" of studying science. This effect, coupled with the performance of the experiments by students and the possible text bias, provided evidence to interpret the significant difference in achievement as only a very tentative conclusion. We noted, however, that the control group teachers tried unusually hard during this period. They felt some pressure.

4. The investigator spent more time assisting the classes using the programed materials than he spent with the control groups. The latter were visited when "time permitted."

In conclusion, the pilot study primarily furnished evidence to support more extensive research in this area of science instruction. Information for curriculum revision suggested changes such as the necessity for more clearly written specific directions in the experiments. Also there was a tendency for some students to follow cues rather than to reason if the frames were poorly written, etc. It was with information such as this that members of the Division of Educational Research turned their attention to Phase II.

PHASE II

The Field Testing of a One-Year Course of Study Using Programed Science Materials at the Fourth Grade Level

A. Objectives

The data from the exploratory first phase was sufficiently encouraging that the decision was made to proceed with the second and much more extensive phase of the study. It should be noted that the primary purpose of Phase II was to write and refine through field testing and analysis a complete course of study for the fourth grade using Skinner's small step approach and his theory of cuing with a linear structure employing constructed responses with immediate knowledge of results. The other objectives are outlined below:

1. To determine whether the performance individually by students of their own experiments remained feasible during the year with increasing spread among the students with time and the consequent necessity for the presence of a multiplicity of laboratory set-ups in the classroom.
2. To measure the achievement of students using programed materials as compared with the achievement of students using a more conventional approach.
3. To ascertain the extent to which students proceeded at varying rates during an entire academic year.
4. To ascertain an expression of the extent of teacher satisfaction with such an approach over a sustained period of time.
6. To determine the validity and level of difficulty of test items; and (1) to establish a relationship between factual achievement and the related concept achievement, (2) to establish the relationship between concept achievement and general science achievement, the results of the analyses to be utilized in revision.

B. Procedure

During the summer and early fall of 1962, units of work were programed in: (1) Heat, (2) Light, (3) Sound, (4) Scientific Method, (5) Plants and Animals, (6) Land, Water, and Air, (7) The Earth and the Sun, (8) Weather, (9) Energy, (10) Friction, (11) Simple Machines, (12) Magnetism and Electricity, and (13) Plants and Animals

Living Together. Most of these materials were administered to two students individually and revised accordingly before the field testing began.

Six fourth-grade classes in four rural elementary schools participated in Phase II. The selection procedures were not random, however, for one teacher was uneasy about the new materials and wished to continue as before, using the programed materials as a guide; a second teacher was a former high school teacher enthusiastic about science who had a gifted group and wished to push her children faster than three thirty-minute periods a week (results from this group are analyzed independently). After some discussion with the teachers and principals, the following decisions were made:

Gordonsville Elementary School.--One fourth grade class to serve as a control.

Unionville Elementary School.--Two teachers to have experimental groups each--one of them a teacher who had participated in the spring pilot study and one who had been a control in the pilot study and now insisted on having the other materials.

Orange Elementary School.--The former high school teacher to have the experimental group but to proceed for five thirty-minute periods a week instead of three and the other teacher to have a control group. A third class was not to participate because many could not read--they were the lowest one-third of the students in Orange fourth grade--academically.

Barboursville Elementary School.--One experimental group.

The above is mentioned to emphasize a limitation to be placed on the conclusions. Since, however, the major purpose of Phase II was to develop a course of study, it would be possible to gather data for curriculum revision and to watch for difficulties that might arise with implications for the design of Phase III.

In late September 1962, the Sequential Test of Educational Progress, Science Test, Form 4a, was administered to students in each of the six classes. Form B was administered in late May 1963.

C. Findings

Descriptive Data. The STEP Science Achievement Test was selected as the criterion measure because it is designed to measure ability to use scientific knowledge to solve problems. The approximate percentages of questions assigned to each aspect of those abilities are as follows:

Define problems	10%	Draw conclusions	23%
Screen hypotheses	25%	Evaluate claims	12%
Select procedures	17%	Reason quantitatively	13%

The distribution of the questions among the six subject-matter fields is: Biology, 40%; Chemistry, 16%; Physics, 23%; Astronomy, 8%; Geology, 7%; Meteorology, 6%.

Forms 4A and 4B are stated to be appropriate for Grades 4, 5, and 6. Each test booklet contains 60 multiple choice items. A single score based on "number right" is obtained. TABLE II presents the initial (Form 4A) and final (Form 4B) mean science achievement based on the "number right" for the experimental group (3 classes), the control group (2 classes) and the one accelerated-experimental class.

TABLE II

INITIAL AND FINAL SCIENCE MEAN ACHIEVEMENT OF THE EXPERIMENTAL, CONTROL, AND ACCELERATED GROUPS

	Form A Initial Mean	Form B Final Mean
Experimental	29	35
Control	25	34
Accelerated	41	42

The number right can be converted into a percentile rank for comparison with the national norms. The percentile rank is used to describe the relative standing of the group or class with respect to the national standing based on a specific comparison group called a "norms group." The "norms group" used to develop the STEP Test consisted of 53 schools (grade 4) and 47 schools (grade 5). The norms are based on fall testing.

TABLE III gives the percentile ranks of the initial and final mean achievement scores of the experimental and control groups, and the accelerated class. The initial means were converted into percentile ranks by comparing the three means of this study with the national percentile distribution of fourth-grade classes, fall testing. The final means were converted into two percentile ranks by comparison with two distributions: (1) fourth-grade classes, fall testing, and (2) fifth-grade classes, fall testing.

TABLE III

INITIAL AND FINAL MEAN ACHIEVEMENT CONVERTED TO PERCENTILE RANKS: NORMS GROUP BASED ON FALL TESTING, GRADES 4 and 5

	Initial Percentile Rank	Final Percentile Rank	
	Grade 4	Grade 4	Grade 5
Experimental	79	94	70
Control	66	92	64
Accelerated	98	98	98

It can be noted from TABLE III that the experimental, control and accelerated classes compare favorably with the national "norms group." For example, 79 per cent of the schools which constitute the national "norms group" showed 4th-grade science means lower than the experimental group on the initial testing; 66 per cent of the schools showed lower 4th-grade science means than did the initial mean of the control group, and 98 per cent of the schools showed lower means than did the accelerated class.

When comparing the final means with the fourth-grade distribution, fall testing, now 94 per cent of the school means fall below the experimental group; 92 per cent, below the control group; and 98 per cent, below the accelerated class. Even when the final means are compared with fifth graders (fall testing), all groups fall above the median (percentile rank = 50).

The percentile rank for the accelerated class is 98 when compared to both the fourth and fifth grade (fall testing) "norms group." The fact that on the initial testing this class reached the ceiling of the test, despite the test's being designed for the fourth, fifth and sixth grades, suggests a limitation of the instrument. Forms 4A and 4B may be more useful for average groups than for groups of a higher ability.

Data Analysis - Achievement. It can be observed from TABLE II that there is a difference of 4 points (29-25) between the initial mean achievement of the experimental and control groups. One criterion analyses of variance were used to determine if not only the initial mean but final mean differences were significant or probably attributable to chance or sampling fluctuations. The analysis of variance summary table is presented below, where the X-scores refer to the initial achievement and the Y-scores to the final achievement.

TABLE IV

ANALYSIS OF VARIANCE OF X AND Y SCORES TAKEN SEPARATELY

Source of Variation	df	SS_x	SS_y	$MS_x(V_x)$	$MS_y(V_y)$
Between Means	1	759	41	759	41
Within Groups	145	13,972	14,881	96.4	102.6
	146				

$$F_x = \frac{759}{96.4} = 7.87^{**}$$

$$F_y = \frac{41}{102.6} = .40$$

**Significant beyond .01 level.

The F-ratio, $F_x = 7.87$ is significant beyond the .01 level, or the probability is 99 times out of 100 there is a significant difference in initial mean achievement between the experimental and control groups.

The F-ratio, $F_y = .40$ is not significant at the .05 level, or the chances are that the difference of one point is due to sampling fluctuations.

The one criterion analysis of variance indicated that a significant difference existed at the .01 level between the initial mean science achievement of the experimental and control groups, the experimental group having a mean achievement = 29; the control, an initial mean science achievement = 25. Because this initial difference may have influenced the comparison of the final mean achievement between the two groups, an analysis of covariance was run to "take out" or hold constant the initial scores, and then compare the final mean achievement between the experimental and control groups. An analysis of co-variance summary table is presented below:

TABLE V

ANALYSIS OF CO-VARIANCE SUMMARY TABLE

Source of Variation	df	$SS_{y.x}$	$MS_{y.x}(V_{y.x})$
Between Means	1	183	183
Within Groups	144	8,158	56.7
Total	145		

$F_{y.x} = 3.23$

The F-ratio does not reach the .05 level or .01 level of significance. Therefore, the null hypothesis is accepted, that after initial effects have been held constant, there is no significant mean difference between the final mean achievement of the experimental and control groups.

Accelerated Class. A comparison of the initial mean achievement and final mean achievement of the accelerated class was done by using the t test for correlated groups, where

$$t = \frac{D}{SE_D} = \frac{M_1 - M_2}{\sqrt{SE_{M_1}{}^2 + SE_{M_2}{}^2 - 2rSE_{M_1}SE_{M_2}}}$$

$$= \frac{41.14 - 42.70}{\sqrt{1.447 + 1.232 - 2(.704)(1.203)(1.11)}}$$

$$= \frac{1.58}{.282} = 5.60, \text{ significant beyond .01 level.}$$

Although the accelerated class gained only 1.58 mean points (as measured by the STEP science achievement test,) the sampling variability was small enough to reject the null hypothesis at the .01 level and declare a significant mean gain in achievement.

Attitudes. In May, 1962, the students and teachers were asked to express their attitudes concerning the science materials. The students' comments remained anonymous, but the school and class were recorded. First, each student was asked to indicate the extent of his like or dislike of the science course of study by selecting one of the options below:

(1) I hate this way of studying science
(2) I don't like this way of studying science
(3) I don't know if I like it or not
(4) I like it a little
(5) I like this way of studying science very much.

The mean like-dislike for the experimental and control groups and the accelerated class were obtained by assigning a value of "1" to the response, I hate this way of studying science, a value of "2" to the next measure of dislike, up to a value of "5" for the response, I like this way of studying science very much. The three measures of mean attitudes are presented as follows:

Experimental group	4.47
Control group	4.57
Accelerated class	4.66.

In general, the mean response of the three groups fell between, "I like it a little" and "I like this new way of studying science very much."

The distribution of the responses of the experimental and control groups are presented in TABLE VI below; e.g., 43 students in the experimental group indicated a 5-response (or they liked it very much) and 49 students in the control group indicated a similar response.

TABLE VI

DISTRIBUTION OF THE RESPONSES, DISLIKE-LIKE OF THE STUDENTS IN THE EXPERIMENTAL AND CONTROL GROUPS

	Dislike----------				----------Like	
	1	2	3	4	5	Sum
Experimental	0	3	4	20	43	70
Control	5	0	0	6	49	60
Sum	5	3	4	26	92	130

In order to determine if the two groups differed significantly in the distribution of responses, a chi square test of independence was calculated; and a chi square value = 19.2 rejected the null hypothesis (at the .01 level), that no relationship exists between the groups and the attitude response (Probability for chi square value at .01 level = 13.27). In other words, the two groups differed significantly in the distribution of attitudes. TABLE VI indicates, for example, that a greater proportion of the students in the experimental group than in the control group make a

4-response. Also five students in the control group selected response 1, while no students in the experimental group expressed a dislike at that extreme. However, a greater proportion of the control group (82%) made a 5-response than in the experimental group (62%).

While the majority of the responses in the experimental group were at the like extreme (5), most of the responses indicated by 4 were made by the students in one class. A chi square test was run against this class and the other two experimental classes. A chi square value = 51.8 was significant beyond the .01 level, which indicates that although the programed materials offer the same content to all classes, systematic effects operate which can cause significant deviation in attitudes among the experimental classes. In summary, although the distribution of responses was different for control and experimental groups, the proportions making either 4 or 5 responses were almost identical, 90% and 92% respectively, for the experimental and control groups.

Expressions of Like and Dislike. In addition to asking the students to indicate their attitude towards science, the investigators requested them to express some of the reasons for disliking the study of science which were given by all the children using the programed materials. (The data are also presented to show the general measure of like or dislike, 1---5.) .

The expressions of like and dislike indicate that the pupils studying science by programed instruction had a generally favorable reaction. Students liked the method of presentation, in particular the experiments and the opportunity to proceed at their own rates. The personal reactions of "interesting," "exciting," or "fun" seemed to permeate their attitudes about science in general.

Most of those students who expressed uncertainty or dislike for the program (responses 2 or 3) also found the program too difficult. Further experimentation should reveal whether if the vocabulary level is lowered and if the method of presentation is simplified, the student who is reading below his average grade level will benefit and will find it a more interesting experience.

The students in the control group likewise had a generally favorable reaction to their science classes. TABLE VIII summarizes their expressions of like and dislike on the attitude inventory. The patterns are similar except that the control group is more pronounced in definitely liking what it likes; i.e., most responses fall into the 5 category - relatively few into the 4.

TABLE VII

DISTRIBUTION OF DISLIKE-LIKE RESPONSE VERSUS REASONS FOR LIKING AND DISLIKING GIVEN BY THE EXPERIMENTAL GROUPS AND ACCELERATED CLASS

	Dislike------Like				
	1	2	3	4	5
Reasons For Liking					
Learning Science					
Like science, educational, new words, good to know, learn more, new things	0	0	2	26	28
Method of Presentation					
Like book					4
Understand better, explains well, easy				2	8
Like experiments	0	0	3	20	42
Own speed				12	12
Do it yourself				3	8
Special subjects				3	6
Personal Reactions					
Interesting, exciting, fun	0	0	0	8	25
Other					10
TOTAL			5	74	143 = 222
Reasons for Disliking					
Learning Science					
Too much science, too long	0	0	0		5
Method of Presentation					
Too difficult, vocabulary too difficult		5	1	11	3
Unclear pictures		5	1	8	9
Tests		1	0	0	9
Too many experiments, lack of equipment		3	1	4	1
Look up words		0	0	0	4
Personal Reactions					
Dull		0	1	0	2
Cheating		2	1	0	0
TOTAL		16	5	23	33 = 77

TABLE VIII

DISTRIBUTION OF DISLIKE-LIKE RESPONSES VERSUS REASONS FOR LIKING AND DISLIKING GIVEN BY THE CONTROL GROUP

	Dislike				Like	
	1	2	3	4	5	
Reasons For Liking						
Learning Science						
Like subject matter, vocabulary, helps a lot, worth doing				3	50	
Method of Presentation						
Learn a lot, new things				4	24	
Easy				1	21	
Experiments				4	21	
Like pictures, simple tests, special subjects				1	7	
Little homework				4	10	
Personal Reaction						
Fun, interesting				4	43	
Other					4	
TOTAL				21	180	= 201
Reasons for Disliking						
Vocabulary hard	5					
Tests	3					
Other	2					
TOTAL	10					= 10

Teacher Attitudes. Many of the criticisms expressed by the teachers using the programed materials are consistent with their students' reactions. Concerning the science content, two expressions of like were the wide variety of subject matter and the organization of concepts. Expressions of like, relevant to the presentation, were favorable to the experiments: "Experiments were many and at the time simple enough for the children to handle without teacher guidance," "Children became familiar with equipment, usually assigned to children in a higher degree of learning." Positive reactions were given regarding the individual rates of speed; e.g., "Individual continuous active participation," "Outstanding pupils advance at a rapid rate of speed, and are not subject to the boredom that is often found when hampered in a slow moving group." The planning of the program and the fact that the "child could find and correct his mistakes," or immediate reinforcement were also mentioned as favorable points of the program.

Two major sources of negative criticisms were apparent. Concern for the below average reader was reflected in suggestions to "simplify the vocabulary," "some of the children will not complete and be exposed to all of the program material," and "weaker children who are exposed to group discussions and explanations gain some knowledge....little place for a poor reader in programed teaching." A second source of criticism was primarily of an administrative nature; i.e., need for space, supplies were sometimes unavailable, lack of clarity of some drawings and some text errors were stressed as sources of concern. One teacher felt that "self expression, creative thinking, and research have little part in programed teaching." Another wished she had more opportunity to participate actively in the teaching.

The control teachers indicated that the written and experimental materials provided by the Division had resulted in a greater amount of science being taught than had been the case in previous years. In some instances they had *learned science* along with the students. They felt that the organization, content and experimental materials were beneficial to the students, although some of the vocabulary was too difficult or extensive for students with lower ability.

D. Conclusions

After the use of programed materials for one year, the mean achievement of the experimental and control groups did not differ significantly. In general, children working independently with programed materials performed as well as those traditional classes which were given assistance above and beyond what is typically received. Any "halo" effect of the programed materials is not a severe limitation because of the one year's duration of the experiment. In addition to any "halo" effects of programed materials, one must likewise consider that control teachers were also working under the stimulus of an experimental program.

The STEP Science Achievement Test purports not to measure facts but the ability to use scientific knowledge. Insofar as this assumption is correct, one can infer that the programed materials did not merely teach factual information, since the final mean achievement of both the experimental and control groups compared favorably with national norms. The materials proved successful in use either as a guide for a course of study for the fourth grade (control group) or as a textbook to be used independently by each child. This result also suggests further investigation into the specific influence of the laboratory-type experiences.

Student attitudes were generally favorable to the program. The students liked the experiments, and in the experimental situation, the opportunity to proceed at one's own rate of speed. The personal reactions of interesting, fun, exciting, seemed to permeate the notion that science is worthwhile, good to know, or an expressed like of the study of science. It is also encouraging that the attitudes of both groups were very favorable; hence, an obvious bias of attitude did not place a limitation on the comparative achievement data.

The teachers expressed a favorable attitude towards the conceptual organization and content of the materials. The performance of the experiments and the variation in rates of progress was administratively feasible. They felt, as did the students who found the vocabulary or reading level "too hard," that some portions of the materials should be simplified. Last, that the teacher is a significant variable was readily apparent. How and why she affects the students in this approach to teaching science are areas of needed research.

In addition to the above findings of Phase II, three related studies were conducted by graduate students as follows:

"Gains in Achievement and Retention of Programed Materials"
by William A. Rodgers

"The Readability of Elementary Programed Science Materials"
by Alton Lee Taylor

"The Prediction of Success in the Fourth-Grade Science Materials"
by William A. Rodgers

These studies originated in attempts to secure additional information from the data that had been obtained. They are not reported here for, although of interest, they are extrinsic to the major purpose of this report. The entire report of Phases I and II and the three mentioned studies may be obtained at cost from the Division of Educational Research at the University of Virginia while they last.

CHAPTER II
PHASE III
THE EVALUATION OF THREE METHODS OF SCIENCE INSTRUCTION AT THE FOURTH GRADE LEVEL

Earlier, several units of materials had been prepared and pilot tested over a one month period. Subsequently, these units were revised and additional units prepared and field tested during the period of one academic year with the science being taught three days a week for one half hour each of the three days. These materials were then revised again based on observations of the students, conferences with the teachers, error counts of the programed materials and analysis of the test data.

A. Objectives

The general purpose of Phase III was to compare the effectiveness of three major methods for the teaching of elementary science at the fourth grade level over the period of one semester: (a) use of a sequence of programed science materials coupled with performance, individually or in pairs, of simple laboratory experiments by the children at such times as they completed certain framed sections; (b) use of these same programed materials by other students who only read about the experiments, but did not perform them, and (c) use of these same programed materials rewritten in textbook format with assignments fixed by the teachers and with periodic teacher demonstrations of the experiments in an emulation of the more typical textbook approach to the teaching of elementary school science. A continuing problem was the revision of the sequential presentation and level of content by empirical means.

The specific objectives were as follows:

Instructional Effectiveness.

1. Achievement. How does the mean science achievement of the three groups of students compare when measured by
   (a) the Sequential Test of Educational Progress (STEP),
   (b) a 200 item T-F test designed and validated by the Research Division?
2. Interest and Attitudes. How do interests of the students in science *per se* compare among the three groups as judged by
   (a) a periodically administered interest inventory, anonymous by student but known by class, and
   (b) amounts of self-initiated science type activities carried on by the students outside of class.
3. Laboratory-type Experiences.
   (a) Is there a difference in science achievement resulting from the use of the laboratory experiences and, if so, is this difference in science achievement seemingly influenced by the interest variable?

(b) How do the three groups compare in their ability to solve problems and to generalize by deductive and inductive type processes?

4. Retention. How does the mean science achievement of the three groups compare when retested at the end of a three month period of summer vacation?

B. Procedure

The field testing took place during the spring semester of 1963 (February through May). The Charlottesville, Virginia, population of 19 fourth grade classes located in seven elementary schools were randomly assigned to each of the three treatment groups by stratified random sampling techniques where the strata were intelligence. Lorge-Thorndike intelligence scores were available for this purpose from the state testing program. The reader should note that it was the custom of the city to group students by achievement and intelligence into two or three levels where the student population in a particular school was large enough to make this possible.

1. Achievement. Initial and final measures of science achievement were gathered in January and June using the STEP Forms 4A and 4B, respectively. The STEP was selected as a measure of student ability to apply scientific principles. The Division T-F test was not

administered initially, but only in May and again in September following summer vacation.

2. Interest and Attitudes. The California Interest Inventory, modified to fourth grade level, was administered at the conclusion of the experiment in May. It furnished two indices: (a) level of student interest in science ranging from practical to more abstract, and (b) type of student interest in science.

   Student attitudes toward the method of instruction were gathered roughly at six week intervals by means of an attitude inventory constructed by the Research Division. These were anonymous by class; i.e., no particular student's attitude could be identified. It was felt, by knowing this, the students would be much more frank in their appraisal.

   To ascertain the extent of student initiated outside activity, straightforward queries were directed to students during the semester.*

3. Laboratory Experiences. Eleven laboratory type problems constituted the initial laboratory resourcefulness test.** For example, one task requested the students, using only a yard length piece of string and scissors, to cut a length of string six inches long. One-half

---

* Limited response to these queries and the difficulty in ascertaining assignments vs. voluntary work, suggested no further data analysis.

**See Appendix A for the other Laboratory Resourcefulness tasks.

the students in each of the three groups were randomly assigned to read about the problems and write solutions. The remainder of the students were provided with simple equipment and an opportunity to manipulate it and then write a solution. These tests were administered individually and, in case of low verbal ability, the student's behavior was observed and his responses recorded.

4. Retention. Science achievement as measured by the STEP Science Achievement and Division True-False Test was evaluated at the end of the usual three month period of summer vacation.

C. Findings

Initial mean differences between the three groups on science achievement and intelligence were tested by the use of a one criterion analysis of variance design. An F-ratio reaching the .05 level of significance rejects the null hypothesis of no difference. These data are presented in TABLES IX and X below where the treatments are defined as:

T_1 = Programed with Student Performed Experiments

T_2 = Programed with Students only Reading about Experiments.

T_3 = Programed in Textbook Format with Teacher Demonstrated Experiments.

1. Initial Measures

a. Initial Measures of Intelligence (Lorge-Thorndike)

	Treatment 1	Treatment 2	Treatment 3
Mean I.Q.	98.9	95.8	98.2

TABLE IX

ANALYSIS OF VARIANCE SUMMARY TABLE
FOR INTELLIGENCE

Source of Variation	df	Sum of Squares	Mean Square (Variance)	F
Between	2	806	403	.93
Within	474	201,525	425	
Total	476			

The F-ratio of .93 was not significant at the .05 level meaning that the hypothesis was accepted that no significant mean difference in intelligence existed between the three groups.

b. Initial Measures of Science Achievement (STEP, Form 4A)

	Treatment 1	Treatment 2	Treatment 3
Mean STEP	31.7	33.3	31.1

TABLE X

ANALYSIS OF VARIANCE SUMMARY TABLE FOR STEP FORM 4A

Source of Variation	df	Sums of Squares	Mean Square (Variance)	F
Between	2	385	192.5	1.3
Within	458	67,454	147.3	
Total	460			

The F-ratio of 1.3 was not significant at the .05 level meaning the hypothesis of no initial mean difference in science achievement was accepted.

It is evident from the above that with respect to intelligence and initial achievement the three treatment groups did not differ significantly. Interestingly enough, however, the six classes of students constituting the second treatment group had the lowest mean I.Q. (95.8) but the highest level of achievement (33.3).

2. Final Measures

a. Analysis of Post-Experimental Test Results in Achievement.

Two criteria analysis of variance designs were used to test differences in measures of science achievement where treatments and levels were defined as follows:*

Treatments: T_1 Programed with Student Performed Experiments
T_2 Programed with Students only Reading about Experiments
T_3 Programed in Textbook Format with Teacher Demonstrated Experiments

Levels: L_1 116 I.Q. and above
L_2 84-115 I.Q.
L_3 Below 84 I.Q.

An F-ratio reaching the .05 level rejects the null hypotheses that (a) no mean differences exist between the three groups in science achievement, (b) no mean differences exist between the three I.Q. levels, and (c) no interaction effects exist between treatments and levels. The means by treatment and level for the criterion measure of STEP in science achievement are reported below.

1. Final Measures of Science Achievement (STEP 4B)

	Treatment 1	Treatment 2	Treatment 3
STEP Means	31.6	31.6	31.9

	Level 1	Level 2	Level 3
STEP Means	47.2	31.6	17.1

---

*Throughout the remainder of this study, reference to treatments and levels is as defined above.

TABLE XI

TWO CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE FOR TREATMENTS BY LEVELS (I.Q.) ON STEP 4B SCIENCE ACHIEVEMENT IN MAY

Source	df	SS	MS(V)	F
Treatments	2	10	5	
Levels	2	40,227	20,113.5	281.3**
Interaction	4	186	46.5	
Within	459	32,832	71.5	

The F-ratio of 295.5 was significant at the .01 level meaning the hypothesis of no mean differences between the three levels in science achievement was rejected. No significant mean differences between the three treatments and no significant interaction effects were found at the .05 level of significance.

2. Final Measure of Science Achievement (Division True-False Test)

	Treatment 1	Treatment 2	Treatment 3
Division T-F Means	137.1	139.1	139.1

**Two asterisks appearing beside a number indicate significance at the .01 level. One asterisk indicates significance at the .05 level.

TABLE XII

ONE CRITERION ANALYSIS OF VARIANCE SUMMARY TABLE FOR DIVISION TRUE-FALSE POST-TEST IN MAY

Source	df	SS	MS(V)	F
Among	2	407	203.5	.36
Within	447	250,511	560.4	

The F-ratio of .36 was not significant at the .05 level meaning the hypothesis of no difference in final mean achievement between the three treatments in science as measured by the Division True-False test is accepted.

b. Type and Level of Interest.

Two criteria analysis of variance designs were used to test differences in type of interest in science and level of interest in science. The type of interest in science is described as a modification of the California Interest Inventory where type of interest in science is defined as meaning the variables, verbal, manipulative, and computational.

The same three hypotheses (see page 40) concerned no mean differences between treatments and none between I.Q. levels and no interaction effects between treatments and levels were made.

1. Type of Interest in Science.

Type of Science Interest by Treatments

	Treatment 1	Treatment 2	Treatment 3
Means	46.5	41.4	42.7

Type of Science Interest by I.Q. Levels

	Level 1	Level 2	Level 3
Means	58.4	43.1	31.6

TABLE XIII

TWO CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE FOR TREATMENTS BY LEVELS (I.Q.) ON TYPE OF INTEREST IN SCIENCE

Source	df	SS	MS(V)	F
Treatments	2	2,180	1,090	1.98
Levels	2	32,091	16,046	29.24**
Interaction	4	2,311	577.7	1.05
Within	456	250,192	548.7	

The F-ratio of 29.24 was significant at the .01 level meaning the hypothesis of no significant difference in type of science by level of I.Q. was rejected. The null hypotheses of no mean difference in level of science interest between groups and of no interaction effects between treatments and levels were both accepted.

2. Level of Interest in Science.

As defined by the California Interest Inventory, level of interest identifies the levels on which an individual's basic interest exists (routine tasks, skilled tasks, professional tasks.)

Level of Science Interest by Treatments

	Treatment 1	Treatment 2	Treatment 3
Means	45.2	40.7	38.8

Level of Science Interest by I.Q. Levels

	Level 1	Level 2	Level 3
Means	52.0	40.4	35.0

TABLE XIV

TWO CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE
FOR TREATMENTS BY LEVELS (I.Q.) ON
LEVEL OF INTEREST IN SCIENCE

Source	df	SS	MS(V)	F
Treatments	2	2,454	1,227	2.2
Levels	2	13,120	6,560	11.7**
Interaction	4	2,278	569.5	1.0
Within	447	251,191	561.9	

The F-ratio of 11.7 was significant at the .01 level meaning the hypothesis of no significant difference in level of science interest by levels of I.Q. was rejected. The null hypotheses of no mean difference in level of science interest between groups and of no interaction effects between treatments and levels were both accepted.

3. Student Attitudes.

Measures of Attitude
Where Values of 1-5 Move from Dislike to Like

	Mean Values		
	Programed With	Programed Without	Text
First Attitude Measure	4.6	4.2	4.8
Final Attitude Measure	4.2	3.8	4.6

Chi square analyses were used to determine any relationships between the measure of like (measured on a 5-point scale) and (1) the three treatment groups, (2) the time spent between the first and last attitude, (3) the progress of the students using the programed texts, (4) the classes within a treatment.

1. The distribution of the first attitude measure vs. the three treatments - chi square value = 51.5 was significant beyond the .01 level of significance. The null hypothesis of independent relationship was rejected between the measure of like and a treatment group. Both Programed With Experiments and Textbook distributions had

greater observed than expected values in the 5-cell (measure of most extreme like), while the Programed Without had a lesser number of observed frequencies (94) than expected (130) in the 5-cell. In other words, more pupils in the Text and Programed With groups expressed an extreme like of the program than was to be expected, while a lesser number expressed this attitude in the Programed Without Experiments group.

2. The distribution of the final attitude measure vs. the three treatments - chi square value = 66.7 was significant beyond the .01 level and the null hypothesis that the treatment groups and the measure of like are independent was rejected. The Programed Without Experiments distribution showed the tendency to have a greater number of observed values than expected in the dislike range, and a lesser number of observed values than expected in the like range (4-5). The Textbook distribution indicated the reverse situation, a greater number of observed frequencies fell in the 5-cell than were expected (118 observed; 94 expected). The Programed With Experiments observed and expected distributions were nearly the same.

The chi square values are consistent with the mean values presented above, which show that both the Textbook and Programed With groups are more enthusiastic initially and finally than is the Programed Without group. However, the final attitude of the Programed With group drops slightly more than that indicated by the Textbook group. One explanation of this is the greater "halo" effect of expectation aroused in the students using the Programed With and Programed Without groups because of the greater departure from traditional methods as compared with the Textbook group.

The mean values of the three groups drop from the first attitude to the last attitude. The following three chi square analyses compare the first and last attitude distribution of each group to determine if a significant change did in fact occur.

3. The distribution of the first and last attitude measures for the Programed With Experiments group (chi square analysis shows value = 19.2) was significant beyond the .01 level. Significant change in distribution did occur for the Programed With group from the first to the last attitude measure. In general, the shift was to less enthusiastic, where the first attitude showed a greater observed frequency (135) than expected (119) in the 5-cell, and the last attitude showed a lesser observed frequency (109) than expected (125) in the 5-cell. Most of the shift on the last attitude moved from the 5-cell to the 4-cell.

4. The distribution of the first and last attitude measures for the Programed Without Experiments group (chi

square value = 10.0) was significant beyond the .01 level. The distribution of first attitudes indicates a greater number of observed frequencies than expected in like range (5-cell) and in the neutral range (3-cell), and a somewhat lesser number of observed values in the 1, 2, and 4-cells. The final attitude distribution shows a tendency for the distribution to shift to the dislike range, where a lesser number of observed values (86) than expected values (90.5) are indicated in the 5-range and a greater number of observed values are indicated in the 1 and 2 categories.

In general the difference between the changes evidenced by the two groups using the programed materials is that the Programed With Experiments dropped frequencies from the 5-cell to the 4-cell, while the frequency distribution for the Programed Without Experiments shifted somewhat to the dislike range (1 and 2-cells).

5. The distribution of the first and last attitude measures for the Textbook group (chi square value = 5.5) was not significant at the .05 level. In general, the distribution for the Textbook group does not change from the first to the last attitude measure.

6. In a consideration of the Programed With group, a chi square analysis was run between the rate of progress of the students and their attitude distribution, - chi square value = 34 rejected the null hypothesis of independence at

the .01 level of significance and declared a relationship between the progress of students and their attitude. In general, the greater the progress, the greater the expression of like.

7. Not only is the influence of progress (and indirect measures relating to progress; e.g., the I.Q. of the student) a significant variable in the attitude formation of students, but also the influence of the teacher or systematic influence on a class should be considered. A chi square analysis was run between the distributions of final attitude vs. the six classes using the programed materials with experiments. A chi square value = 64.9 rejected the null hypothesis, which indicates that attitudes of classes do vary within this treatment group.

A consideration of the changes of the distribution from the first attitude measures to the last measures should be established for each teacher within a group (when chi square values show differences in the attitude distribution among classes within a given group). These data must then be compared with the rate of progress, and the ability of the class.

c. Laboratory Experiences.

An attempt was made to evaluate the effects of the laboratory experiences by means of a laboratory resourcefulness test constructed by the Research Division. A three criteria analysis of variance design was used to test differences in mean achievement where

A = The Three Treatments
B = The Three I.Q. Levels
C = Students Read about vs. Manipulated Simple Science Equipment

Means on Laboratory Tests by Treatments

	Treatment 1	Treatment 2	Treatment 3
Means	17.1	16.0	16.4

TABLE XV

MEANS BY TREATMENTS AND LEVELS ON LABORATORY RESOURCEFULNESS TEST

I.Q.	Treatment 1	Treatment 2	Treatment 3
Level 1	23	18	24
Level 2	16	16	16
Level 3	14	13	13

TABLE XVI

THREE CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE FOR LABORATORY RESOURCEFULNESS TEST

Source	df	SS	MS(V)	F
A	2	83	41.5	3.80*
B	2	2207	1103.5	101.23**
C	1	1	1	.09
AB	4	222	55.5	5.09**
AC	2	55	27.5	2.52
BC	2	70	35.0	3.21
ABC	4	76	19.0	1.74
Within	378	4121	10.9	

The F-ratio of 3.80 was significant at the .05 level meaning the hypothesis of no significant difference in laboratory performance by treatments was rejected. Thus, the laboratory resourcefulness test reflected student improvement in ability to solve new types of problems using simple equipment.

A significant interaction effect was found between treatments and intelligence levels where a larger mean difference in performance was found at the highest I.Q. levels.

d. Retention.

Science achievement was assessed following the three months of summer vacation using the STEP Science Achievement Test Form 4B and the 200 item Division constructed True-False Test. One criterion analysis of variance designs were used to test differences in mean achievement as measured by these tests.

1. Final Measure (Following Summer Vacation) of Science Achievement - STEP 4B

	Treatment 1	Treatment 2	Treatment 3
STEP Means	32.7	33.6	33.2

TABLE XVII

ONE CRITERION ANALYSIS OF VARIANCE SUMMARY TABLE FOR STEP 4B RETENTION--FOLLOWING SUMMER VACATION

Source	df	SS	MS(V)	F
Among	2	575	287.5	1.8
Within	397	63,307	159.5	

The F-ratio of 1.8 was not significant indicating the hypothesis of no difference in finally retained mean achievement in science was not rejected. One method was as good as another after this three months period. Similar results were secured using the True-False test as follows.

2. Final Measure (Following Summer Vacation) of Science Achievement - Division True-False Test

	Treatment 1	Treatment 2	Treatment 3
Means	138.4	139.6	138.6

TABLE XVIII

ONE CRITERION ANALYSIS OF VARIANCE SUMMARY TABLE FOR DIVISION TRUE-FALSE TEST IN RETENTION FOLLOWING SUMMER VACATION

Source	df	SS	MS(V)	F
Among	2	103	51.5	.08
Within	395	248,667	629.5	

The F-ratio of 1.8 was not significant indicating the hypothesis of no difference in finally retained mean achievement in science was not rejected.

D. Conclusions

It will be recalled that this summary reports the results of using the three major treatments: (1) six classes of students using the programed materials with the experiments, (2) six classes of students using these same materials but only reading about the experiments, and (3) six classes using these materials rewritten in textbook form with the teachers doing the experiments in demonstration.

As indicated, there was no initial significant difference between these treatments with respect to the factors of intelligence, or initial science achievement. In post-testing, it was found that with respect to interest, there was no mean significant difference between the three groups, but that there was a significant difference by levels of intelligence. The lower I.Q. group was least interested, the middle I.Q. group more interested, and the highest I.Q. group most interested. Since each treatment group contained all I.Q. ranges, the factor of either doing the experiments or seeing the teacher do them was not apparently a significant factor.

With respect to type of interest in science; i.e., at what level of abstractness did the various students indicate like for science, again the factor of I.Q. was significant. This is to be expected. Thus, for example, if a student indicated an interest in being engaged in the occupation of a florist, the less able might choose to drive a truck for a florist shop rather than do the bookkeeping, etc.

The significant interaction effect between treatment and I.Q. level in laboratory performance suggests the value of concrete experiences should not be overlooked for the brighter child. In addition, the findings indicate that children who are allowed to actively participate in performing laboratory experiments have a greater proficiency in the solution of simple problems.

With respect to science achievement, as measured by the STEP test, there was no difference between treatments. Again, as would be expected, there was great variation according to ability as measured by intelligence. The same thing was true of science achievement as measured by a 200 item True-False 'fact' test.

When tested at the end of the summer for delayed retention, there was again no significant difference between treatments, whether measured by the STEP Science Achievement Test or by the Division constructed True-False test.

With respect to attitude toward science (as contrasted with the interest measure) the second treatment group disliked science more than either of the other two. This was significant at the .01 level. This was maintained through the experiment's duration of one semester. In short, those two groups having some direct contact with experimentation whether performed by the students themselves or through observation of teacher demonstrations reported the highest degree of liking for science.

In addition, the extent to which these three groups maintained their attitudes varied. The textbook group maintained a consistently high attitude of like for science. The programed group with students doing the experiments shifted to a slightly less positive position of liking--although still positively oriented. The programed group that only read about the experiments shifted to distinct dislike.

A very interesting point in all this is that the variation in attitude was not reflected in proportionate variations in achievement. This is unexplainable at present. The students who liked science the least (as taught) were the group who only read about the experiments, but they achieved just as well. On the other hand, from a chi square analysis within one group (the programed science group doing experiments) there was clearly evident a positive relationship between attitude and achievement. The null hypothesis of independence between attitude and achievement was rejected at the .01 level.

Another puzzling aspect of the experimental results was since students in the two groups using the programed materials were allowed to progress at varying rates, why wasn't this reflected in an increased mean achievement for those groups? No explanation is available for this as yet.

In any event, revisions were made in the materials during the summer in readiness for the extensive program to be begun that fall.

CHAPTER III
PHASE IV
THE INFLUENCE OF THREE RESPONSE FACTORS IN PROGRAMED LEARNING

By the summer of 1963, we had (1) pilot tested several units during the spring of 1962, (2) revised these units, prepared others, and field tested these during the following academic year with the science being taught three days a week for one-half hour each day, and (3) re-revised some of these materials for use during the spring semester of 1963.

In preparation for the fall semester (1963) investigation of the influence of three response factors, there was one further revision of the following units: heat, light, sound, science and its ways, land,water and air, plants and animals, weather, the earth and the sun, plants and animals living together, electricity, and the earth. The remaining units were not revised at this time because the testing period was only for 18 weeks.

One problem that still plagued us was that we had paid too little attention to the reading difficulty level. It was too high for the lowest one-fourth of our students. This was made clear in the analysis using the Dale-Chall Formula[1]. This did not introduce a biasing factor because of the design, but since the classes were grouped by general ability factors such as achievement and intelligence, it necessitated some special arrangements for these lowest ability students (See Appendix B). We attempted to lower the level somewhat in our materials modification, by simplifying sentence complexity, but felt we should not make drastic changes at this stage in the experiment.

---

[1] Taylor, Alton L., "The Readability of Elementary Programed Science Materials," The Field Testing of a One-Year Course of Study Using Programed Science Materials at the Fourth Grade Level, Division of Educational Research Bulletin No. 1, University of Virginia, January, 1964, pp. 32-35.

A. Objectives

The objectives of this study were the same as for the preceding spring except that in addition the influence of three response factors in programed learning were investigated. These objectives consisted of (1) achievement, (2) interest and attitude, (3) the effects of laboratory-type experiences, and (4) retention. Thus, the same measures of science achievement using the STEP, the Division True-False, the usual I.Q. measures, the interest and attitude indices and the observations by the research assistantswere employed. Also, every student would do the experiments whereas before only those in Treatment 1 had done so, while Treatment 2 had allowed the students to read about them, and Treatment 3 had involved teacher demonstrations.

Thus, in addition to the objectives listed and described in detail in Chapter II, three questions were asked concerning the most effective response variables or variable combinations:

1. Constructed response vs. multiple choice
2. All answers vs. partial feedback
3. Branching vs. linear programing.

This necessitated the preparation of eight different versions of the programed materials as follows:

1. Linear - Constructed Response - All Answers
2. Linear - Constructed Response - 1/3 Answers
3. Linear - Multiple Choice - All Answers
4. Linear - Multiple Choice - 1/3 Answers

5. Branching - Constructed Response - All Answers
6. Branching - Constructed Response - 1/3 Answers
7. Branching - Multiple Choice - All Answers
8. Branching - Multiple Choice - 1/3 Answers

In summary, the fact that the programed approach is just as effective in achievement as the textbook approach was now clearly established. The purpose here was to ascertain whether one or several of the above variations in format seem to have more promise than the others.

B. Procedure

The Charlottesville, Virginia school population consisted of 20 fourth grade classes located in seven elementary schools. All eight versions were randomly assigned within each of the 20 classes of students so that each teacher had a few students using each of the eight versions.

About half of the teachers who had worked with us during the spring semester were involved in this phase. The attrition was due to such factors as married teachers whose husbands had completed their work at the University and moved, the shifting of several teachers to other grade levels, retirement, pregnancy, and the addition of one new fourth grade class to the system because of increased enrollment. All teachers who had worked with the materials the previous spring and who remained in the system were willing to participate.

As before, one or more graduate assistants were assigned to each school to assist teachers in (1) administering tests and recording test data, (2) insuring the necessary science equipment would be available as needed, (3) noting administrative or other difficulties that might arise, (4) overseeing experiments that were potentially dangerous such as the lighting of candles, (5) transporting booklets and test data to and from the schools, et al.

C. Findings

Initial mean measures for comparison between the students in each of the eight versions were tested by the use of a one criterion analysis of variance design. The three measures compared were (1) science achievement on the STEP, (2) science achievement as measured by the comprehensive T-F test, and (3) intelligence. An F-ratio reaching the

.05 level of significance rejects the null hypothesis of no difference. These data are presented in TABLES XIX, XX, and XXI where the treatments are defined as:

T_1 = Constructed - Linear - All Answers
T_2 = Constructed - Linear - 1/3 Answers
T_3 = Constructed - Branching - All Answers
T_4 = Constructed - Branching - 1/3 Answers
T_5 = Multiple Choice - Linear - All Answers
T_6 = Multiple Choice - Linear - 1/3 Answers
T_7 = Multiple Choice - Branching - All Answers
T_8 = Multiple Choice - Branching - 1/3 Answers

1. Initial Measures

a. Initial Measures of Intelligence (Lorge-Thorndike)

Treatment 1	100.67	Treatment 5	100.57
Treatment 2	104.20	Treatment 6	103.53
Treatment 3	104.52	Treatment 7	101.61
Treatment 4	100.91	Treatment 8	102.16

TABLE XIX

ANALYSIS OF VARIANCE SUMMARY TABLE
FOR INTELLIGENCE
FOR EIGHT TREATMENTS

Source of Variation	df	Sum of Squares	Mean Square (Variance)	F
Between	7	1,059	151.3	.61
Within	480	119,010	247.9	
Total	487			

The F-ratio of .61 was not significant at the .05 level meaning that the hypothesis was accepted that no significant mean difference in intelligence existed between the eight groups.

b. Initial Measures of Science Achievement (STEP, Form 4A)

Treatment 1	31.11	Treatment 5	30.60
Treatment 2	32.05	Treatment 6	32.84
Treatment 3	34.02	Treatment 7	29.79
Treatment 4	33.61	Treatment 8	31.75

TABLE XX

ANALYSIS OF VARIANCE SUMMARY TABLE
FOR STEP
FOR EIGHT TREATMENTS

Source of Variation	df	Sum of Squares	Mean Square (Variance)	F
Between	7	915	130.7	1.1
Within	485	59,388	122.4	
Total	492			

The F-ratio of 1.1 was not significant at the .05 level meaning that the hypothesis was accepted that no significant mean difference in initial achievement as measured by STEP existed.

c. Initial Measures of Science Achievement (True-False Test)

Treatment 1	127.70	Treatment 5	124.14
Treatment 2	125.16	Treatment 6	126.66
Treatment 3	128.81	Treatment 7	127.10
Treatment 4	127.42	Treatment 8	125.90

TABLE XXI

ANALYSIS OF VARIANCE SUMMARY TABLE FOR TRUE-FALSE TEST FOR EIGHT TREATMENTS

Source of Variation	df	Sum of Squares	Mean Square (Variance)	F
Between	7	723	103.3	.44
Within	486	145,309	299	
Total	493			

The F-ratio of .44 was not significant at the .05 level meaning that the hypothesis was accepted that no significant mean differences in initial achievement as measured by the Division True-False test existed.

It is evident from the above that with respect to intelligence and initial achievement as measured by both the STEP and the Division True-False test, the eight treatment groups did not differ significantly.

2. Final Measures

a. Final Measures of Science Achievement (STEP, Form 4B)

A three criteria analysis of variance design was used to test differences in mean achievement where

A = Type of Response (constructed vs. multiple choice)
B = Type of Answer (all answers vs. one-third answers)
C = Format of Material (linear vs. branching)

The coding is

A_1 = Constructed Response
A_2 = Multiple Choice

B_1 = All Answers Given
B_2 = One Third Answers Given

C_1 = Linear Structure
C_2 = Branching Structure

The three criteria are represented as follows:

$$
\begin{array}{cccccccc}
\multicolumn{4}{c}{A_1} & \multicolumn{4}{c}{A_2} \\
\multicolumn{2}{c}{B_1} & \multicolumn{2}{c}{B_2} & \multicolumn{2}{c}{B_1} & \multicolumn{2}{c}{B_2} \\
C_1 & C_2 & C_1 & C_2 & C_1 & C_2 & C_1 & C_2
\end{array}
$$

The initial and final means for each of the eight versions are presented in Appendix C. The analysis of variance summary tables comparing the three criteria (linear vs. branching, all vs. 1/3 answers, constructed vs. multiple choice) are as follows:

TABLE XXII

THREE CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE FOR FINAL ACHIEVEMENT TEST (STEP)

Source of Variation	df	Sums of Squares	Mean Square (Variance)	F
A	1	8	8	
B	1	0	0	
C	1	1	1	
AB	1	35	35	
AC	1	4	4	
BC	1	110	110	
ABC	1	1	1	
Within	488	56,992	116.8	
Total	495			

The F-ratios are all insignificant meaning the hypothesis of no significant difference in achievement by treatments was not rejected. One version is apparently as effective as the other.

b. Final Measures of Science Achievement (Division True-False Test)

A three criteria analysis of variance design using the same coding as above was used to test the difference in final mean achievement on the Division True-False test.

TABLE XXIII

THREE CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE FOR FINAL ACHIEVEMENT TEST (TRUE-FALSE)

Source of Variation	df	Sums of Squares	Mean Square (Variance)	F
A	1	105	105	
B	1	179	179	
C	1	3	3	
AB	1	24	24	
AC	1	92	92	
BC	1	112	112	
ABC	1	313	313	
Within	488	359,988	737.7	
Total	495			

The F-ratios are all insignificant meaning the hypothesis of no significant difference in achievement by treatments was not rejected. One version is apparently as effective as the other according to the Division True-False test.

c. Final Measure of Level of Interest

A three criteria analysis of variance design using the same coding as above was used to test the differences in final mean achievement on the level of interest in science by the students in the eight treatments.

TABLE XXIV

THREE CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE FOR FINAL LEVEL OF INTEREST

Source of Variation	df	Sums of Squares	Mean Square (Variance)	F
A	1	464	464	
B	1	1378	1378	2.33
C	1	182	182	
AB	1	23	23	
AC	1	2199	2199	3.72
BC	1	472	472	
ABC	1	56	56	
Within	480	283,297	590.2	
Total	487			

Although none of the F-ratios were significant, there was an almost significant value of 3.72. This variance (AC) indicates some interaction effect between type of response and format of material. In particular, the treatment consisting of constructed response in linear form with only one-third of the answers given was the most effective.

d. Final Measure of Type of Interest

TABLE XXV

THREE CRITERIA ANALYSIS OF VARIANCE SUMMARY TABLE FOR FINAL TYPE OF INTEREST

Source of Variation	df	Sums of Squares	Mean Square (Variance)	F
A	1	1,314	1,314	
B	1	79	79	
C	1	2	2	
AB	1	13	13	
AC	1	21	21	
BC	1	3	3	
ABC	1	820	820	
Within	480	626.1	1.3	
Total	487			

The F-ratios are not significant meaning the hypothesis of no significant difference in type of interest by treatments was not rejected. One version is apparently as effective as the other according to the degree to which students are interested in science at the various levels.

D. Conclusions

The influence of three response factors in programed learning was thoroughly investigated. This entailed eight different versions randomly distributed throughout the twenty classes of students. The F-ratios on all achievement indices were not significant. This was also the case with interest factors although there was some interaction effect between type of response and format of material. In particular, the treatment consisting of constructed response in linear form with only one-third of the answers given was the most effective in terms of level of interest.

Despite the many weaknesses evidenced by the division in quality of preparation of materials, faulty experiments, ignorance of knowing how to introduce materials to teachers unfamiliar with them, and the many many unanticipated problems that arise in any such large scale experiment, the prevailing attitude of both teachers and students after the months of working with them was positive. This in itself is an indication that, if materials can be sufficiently well prepared, if teachers can be trained, there is a place for this approach in elementary school science.

E. Subjective Appraisal of the Project

In addition to the test data and the formal attitudinal reactions of students, it was considered desirable to secure written reports of a subjective nature from research assistants, teachers, and students concerning the strengths and weaknesses of programed materials as a means of teaching science in the elementary school. These observations are presented in summary form below.

1. Observations by Teachers

Perhaps the best way to describe the teachers' reactions is to report the summary statements of four of them exactly as they were written.

Teacher A: Excellent for mature superior children. Absolute waste of time for slow learners and immature classes such as mine (with a few exceptions). Glad it is over so they can learn material suitable for their understanding and level of development.

Teacher B: The programed science has been very frustrating as far as this teacher is concerned -- because of the restrictions. However, it is very evident that there are certain children capable of being their own teacher and they have progressed more than satisfactorily. The poor reader has refused to set his own best learning pace. He has raced through -- just marking his answers -- consequently missing three-fourths or more of the questions on the tests.

Teacher C: The material is excellent. It develops good work habits, increases powers of concentration, is good preparation for T-V science and encourages interest in research. The children soon discover that rushing through the materials is not wise.

The approach encourages use of the dictionary, allows each child to progress at his own speed, encourages outside of school interest in science and is inspirational and instructive for the teacher.

The children loved the experiments. For myself I felt it was a privilege being a part of the University program.

Teacher D: Content: The subject matter covered is excellent, though presented in the wrong sequence. Sound, Light, etc. (the harder subjects for 9-10 year olds) should follow the unit "Science and Its Ways" as they should begin teaching children where they are.

Approach: The approach is stimulating to children -- they love the variation of answering questions, seeing answers, doing experiments, etc. Though the approach is a little mature for some pupils, most average to bright ones understand what to do.

Experiments: Here the program could be much improved. The directions are good, but materials are inadequate and experiments often confusing and contradictory. Light and Sound experiments are very poor -- please make them more self explanatory.

Vocabulary: For bright children the vocabulary is fine, but much too difficult for slow to average children.

Testing: At times the questions are vaguely worded. They do better on objective tests, as fill-in-blank questions only test their ability to memorize. The objective tests are the ones which children score highest on, which they enjoy most, and which are easiest to grade.

Results: The tests show that the mature, average to bright children understand a great deal of the material presented. The less mature (about 30%) of any group do not understand the material -- concepts, vocabulary, experiments, etc. These children are too young to study on their own -- they still need a teacher.

Suggestions: You should down grade in difficulty your beginning volumes and begin where the child is in emotional and mental maturity.

These were but four of the reactions from about fifty teachers. It is evident that they were frank and that they were varied. The most frequent single comment was that we failed to reach the slower student -- and in this they were exactly right. We just did not simplify our vocabulary sufficiently.

As a general figure, about 80% of the teachers favored the approach although all felt there was room for improvement in specific respects.

2. Observations by Students.

Perhaps one way to report student reactions is to quote some of their comments:

Negative

1. Some experiments are too hard. Some are too easy.
2. I think some of the words are too hard.
3. We should all circle our answers. I get tired of writing down all the answers in the blank spaces.
4. I don't think the experiments are well-spaced. Some are too close together. Others are too far apart.
5. I don't like some pupils having all of the answers while others only have one third of the answers.
6. I think the tests are too long and too hard - so are the books. I think the books are very dull.
7. I wish they had more stories or adventures.

8. When people look at your book and see that you got a lot of answers wrong they'll say, "Ooh, this kid's a stupid kid."

9. Sometimes I move my paper (covering the answers) down a little too far and then move it back up. One time I did that, and I wanted to be honest about it and before that I thought the answer I was going to put and it was the wrong answer, but I circled it anyway.

10. I daydream half the time, but not as much as I work.

11. When I'm on, say a real hard page, and I'm just trying to figure it out and she (teacher) says, put the books up, then the next day I say, "Oh brother, I wish we didn't have programed science," because then I've got to figure it out. Kinda complicated.

12. I don't understand all of it.

Positive

1. I get to do experiments.

2. I study at my own speed.

3. I learn a lot.

4. The new way is different.

5. I work by myself.

6. It's so good I don't think you could change anything about it.

7. You're on your own honor and you try not to cheat.

8. When we had regular science the teacher would keep putting off the experiments because we didn't have the stuff to do it, but now it's all here.

9. That's neat being able to check your own answers.

10. I like the way we are studying science because we learn science ourselves.

11. You get to go outside.

12. I wish we had a lot more units.

These were the most common types of both negative and positive reactions from students. In all fairness, however, it should be noted that positive reactions far outweighed negative reactions even after a period of three or four months. As an example, an attitude inventory was administered to 92 students in three classes in one of the elementary schools. This occurred on December 18, approximately ten weeks after the start of the experiment. The children's reactions are summarized below:

	Dislike Very Much	Dislike	Like	Like Very Much
Like				
1. I get to do experiments	0	0	13	79
2. I study at my own speed	1	3	16	72
3. I learn about new things	2	3	17	70
4. I learn a lot	1	3	19	69
5. The new way is different	6	5	16	65
6. I work by myself	4	2	25	61
7. I don't have to write so much	3	7	21	61
8. I learn new words	0	4	30	58
9. I teach myself instead of the teacher doing it	7	4	23	58
10. This new way is interesting	1	5	36	50
11. I get the answer at once	4	5	30	53
12. New way is easy and fun	2	8	37	45
Dislike				
1. The new way is very long	9	11	26	46
2. I have to use a dictionary	20	22	20	30
3. I think the words are hard	5	20	38	29
4. I get bored	32	38	9	13
5. I don't understand all of it.	13	38	26	15
Ambiguous Responses				
1. I write my answers in small blanks.	10	14	32	36
2. I get to take the tests	20	12	35	25
3. I can look at my answers before I write them.[2]	42	19	14	17

---

[2] J.F. Crabtree, II. An independent study conducted on programing experiment, Div. of Ed. Res., Univ. of Va., 1963.

The approach used was to ask the students to respond to statements on a four point scale. The statements were derived from the original unstructured most typical comments of children in the earlier study. These children encompassed all three ability levels.

It is evident that the vast majority of students were positively oriented toward this approach ten weeks after it had started.

3. Observations by Research Assistants

Five of the briefer case studies exactly as written by the graduate research assistants are reported below. Each case study refers to a particular class with a particular teacher.

A. The children were average and above in intelligence. From the beginning to the very end of the period the class as a whole progressed very satisfactorily. The teacher knew what was following and was prepared to help the pupils. She not only had the experiments set ahead of time, but she also worked with the students at the experiment table and explained further the results of each experiment. She was prepared to answer and explain to those pupils who had difficulty in comprehending the working of frames or the meaning of words. She questioned to determine if the meaning had been gained from the activities.

The class was always orderly and the pupils were busy with either their experiments or with their books. The test that each pupil took after having completed a certain unit was corrected by his teacher and returned to the pupil to see his mistakes. If he made a satisfactory grade, he was permitted to start on the next unit; if not, he used to review the unit and take another test.

This teacher was the most independent, but asked us to suggest or recommend certain procedures. Often she consulted on the progress of students from unit to unit, the materials presented in the book, and the tests.

The class was very much interested in the program as well as the teacher. This was evidenced by the seriousness with which all pupils were studying the materials and the interest they showed and the excitement around the experiment table. The teacher elected to return to the other science book at the end of the experimental period because of the materials she wished to continue with (more in space flight). However, she continued to utilize the experiment table as was the practice with the experimental materials in the more traditional approach.

B. The pupils in this class were generally average or below in intelligence. They could neither concentrate nor stay in their seats. They spent much time around the experiment table either doing the same experiment over and over again or doing the experiment for somebody else.

The mood of the class varied with that of the teacher. Inconsistency was the "pattern." During certain periods the teacher was not able to do several things, like answering questions, handing out tests, grading them, preparing experiments. However, when she was in a good mood, she could be an extremely competent teacher. She was very much aware if she was not doing an adequate job. The inconsistent behavior probably contributed to the pupils' not wanting to ask the teacher questions, probably because of fear, loss of fruit break, etc. The difference in class attitude, commotion, confusion, indicate the teacher was the key figure in the class.

The teacher seemed to like this program most of the time. The pupils, too, did like it and wanted to continue with it. The class elected to continue with the materials on a part-time basis. A great deal of science interest continued in regular science classes after the termination of the formal experimental period. The teacher was creative to her approach to teaching in general. She did not object to the more structured programed science approach, but had she been more oriented towards science or had more consistent communication with the class, this might have been a problem.

C. Superior intelligence describes this class of pupils. The class continued with the program after the experiment terminated. The teacher is very much interested in the program, and in the theory of programing. She programed units in grammer for students. Her major interest is social studies, but she shows enthusiasm for all phases of learning. It is difficult to determine if enthusiasm all comes from the teacher or what part is inspired by the bright group of children. The usual class procedure was nondirective, so pupils adapted well to programed units. The class was orderly, pupils busy doing work, and accepting responsibility for learning quite readily.

The experiments were set and she was prepared to help them, although often times she turned the responsibility in directing over to us. She assumed the responsibility of administrative tasks; e.g., grading tests, handing out tests. The teacher's competence was demonstrated on things she was sure of, and she derived a great deal of pride from students' work. When in doubt, she delegated responsibility.

Usually the science class did not start on time because the teacher was doing something else before and continued the same. The pupils usually informed her of the time to begin science. The pupils' enthusiasm was evidenced by the very interesting questions they asked covering the material and doing the experiments. Discipline and confusion were never a problem. The class moved fast, smoothly, without any irregularities. The pupils showed initiative in doing things. Some of the materials were below their growth, but students were not insulted. They enjoyed repeating the activities. Their one criticism was repetition in the book on learning to spell certain science words. The class had an efficient but not dictatorial tone.

D. The children I worked with at ________ school were an accelerated group with an average I.Q. of 120. They were unique in so far as they were extremely compatible. Not only were they well disciplined by their teacher but they exhibited excellent control within their group.

The children were highly motivated, very sensitive and responded beautifully. Their parents were mostly professionals.

In the beginning of the program many of the children rushed through the book to get to the experiments. This was to be expected even with this group.

After a period of time, they slowed down and most seemed to find their own rate of speed. Of course there were a few who competed with their neighbors and as a result were careless in their work.

They had definite likes and dislikes. Their chief complaint was the repetition. Their second complaint dealt with the differences between the 1/3 version and the all answers version. Those with only 1/3 of the answers said it wasn't fair. They also didn't like writing in the answers. They preferred circling them in the multiple-choice versions.

They loved the experiments and the idea that they were independent and could do things on their own. Some of the comments made by this group were: "It was fun. It makes you think. We learn new things, learn new words. The sub-tests help prepare you for the final tests."

In the beginning it was a novelty and all were very enthusiastic. Toward the end of the program some became discouraged and others did not feel like exerting the effort needed.

This method of teaching seemed to discourage extra research and study in a particular area. They had the feeling that they must go on in their books and couldn't take the time to do work outside of class.

E. This study is condensed to about one half of original length as written by the research assistant.

Measured I.Q.'s ranged from 76 to 118. There was obviously an even greater range of home environment.

The teacher has little knowledge of science. Her method of teaching is autocratic and she is accustomed to "teaching the children" rather than having them learn. She likes to keep her class organized. She tends to be harsh on the children, scolding them for things she thinks they have done. Children find it hard to approach her with their questions and problems.

With few exceptions, the students are not autonomous. They are accustomed to the educational "lock-step" and to being scolded when they step out of line. Most learning is passive. They just sit and wait to be told what to do.

Initial student reaction was very good. Although they were fearful of doing something wrong as they worked along, and they constantly asked what they should do, they seemed to like the type of course. They especially liked the experiments.

Scores on the first section test ranged from 11 to 21 right out of 25 questions. The teacher seemd rather upset that the children were not learning the material, and I agreed that we should try to slow them down a little. We counseled the class not to rush through too fast. I think the teacher said it a lot more strongly when I was not there.

Work on the second section was much slower. Most students were reluctant to take the section test, preferring to review the material in accordance with the teacher's wishes. The review accomplished little. Test scores were not better than on the first test. The teacher also contributed directly to the slow down by not letting the children take the test when they first asked for it. She also insisted that the tests be corrected in every detail, which took a great deal of time. Because of her harsh manner, the children often delayed taking a paper to her.

The children loved the experiments. They liked the opportunity to do them and they liked the freedom. Many children did them over and over again.

The teacher was very good about having the equipment ready for the children. As we got into sections where she had to get materials, she began to complain about the work and the difficulty of securing them (she is elderly) and I progressively did more and more of it.

Although they enjoyed it, the children did not seem to get much out of the experiments. Their approach was as to toys or to magic. Few could tell me what they had observed in the experiment without my helping them. Some of the experiments worked very poorly, if at all. Some seemed not to have been tried by the person who put them into the program.

In my opinion, much of the testing is not accurate for this class. The low reading skills are a severe handicap. Their desire to provide the right answer leads to a tendency to put down the answer they think the teacher wants and to dishonesty. Many of the children are incapable of autonomous thinking.

Other Comments: Teachers obviously must be trained to use this material. They must see a value to it and be made to realize that children can learn from it. A teacher who believes that children must be taught will feel that she is not doing her job. Her attitudes are conveyed to the children. She must also be taught to be patient while the children adjust to this new method.

Children also have to learn to use it. They must learn to appreciate the printed material as well as the experiments. Better preparation of the material should help a lot.

Some instruction on doing experiments is needed badly. Children must be led to the observation they are to make. Many of the experiments are just instructions to do something. They should be told what they are to look for.

Teachers should review workbooks regularly. One child in this class skipped about half a section and never realized it. I discovered it after he had taken the test and missed most of the questions. Also, many children never ask for help and the teacher should not let them drag along.

Since the children are completely on their own, the teacher has no indication of their progress or understanding until they take the test. I have discovered consistent errors in workbooks that led to incorrect answers on tests. These errors could have been corrected before testing. This is another reason for the teacher to check workbooks.

When all answers are not given, it is easier for a child to make errors that are not corrected. Care should be taken that the answers that are given cover all the material so as to make it impossible for a child to make consistently wrong answers and not know it.

The experiments could be better devised to provide the learning that is desired. Many of them are quite clever but others are not as good as they could be.

Much economy of equipment could be obtained by covering less material and going into it deeper, because some of the equipment could be used on several experiments. The savings might be used to provide better equipment, in some cases, or a few basic implements such as a metal stand and holder, a Bunsen burner or hot plate, a sounding box, a dark box for observing light experiments, a convenient electrical outlet, or a magnetic compass.

Recommendations:

1. Time limits should be placed on tests and children should complete them in one class period.
2. Review should be discouraged. The program itself provides for reinforcement and review.
3. All experiments should be thoroughly pre-tested and all materials that can affect the procedures on the outcome of the experiment should be furnished.
4. All experiment sheets should include questions to be answered during the experiment and data to be gathered if possible. Each experiment should be introduced with a discussion of what is to be done and what is to be observed.
5. Attitude and interest surveys for this class should be discarded.
6. Section test for Section 2, Volume I should be discarded.
7. All written material be revised or rewritten completely.
8. The presentation of refraction of light be omitted or further tested to see if this is understandable to fourth graders.
9. At least two versions of this material be prepared. One version with low reading level for slow students should include much repetition. The other for brighter children should omit some of the repetition and could have a higher reading level.
10. All workbooks be reviewed at least once a week.
11. The number of different topics be reduced considerably. A study in depth of one or two topics would result in better understanding of the topic, a better understanding of the nature of science, better utilization of equipment, and better reinforcement of vocabulary additions.

In addition to the five case studies reported above, all eighteen case studies were examined for two types of comments (1) negative, and (2) positive. These are reported below.

A. Negative Comments as Seen by Research Assistants

Materials:

1. Wrong answer obtained by student because of misinterpretation of material.
2. Vocabulary too difficult for some students.
3. Some of the art work is not clear.
4. Some of the books were sloppily bound - missing pages.
5. Some weaknesses were due to poorly written materials rather than to programing as a methodology.
6. The texts are written as miniature physics courses, etc. Should be more interrelated.
7. Some thought materials were too repetitious -- especially for the bright.
8. The teacher griped loudly about the materials in front of the class.
9. There were inconsistencies in instructions and ambiguity in some of the definitions.
10. Too many concepts were introduced.
11. The material needs thorough revision of content, format, style, and methodology.
12. The reading level is too difficult for poorer readers.

Experiments:

1. Some children loiter over experiments.
2. Certain experiments did not work well.
3. Much of the work at the experiment tables was sheer play.
4. The children did not seem to get much out of the experiments.

Students:

1. Some students go faster than they should -- some slower.
2. Some students waste time at their desks.
3. Some students loiter over experiments.
4. Some students are deficient in reading skills.

5. In the slow class some children pretended to work as if they understood.
6. Some students always needed to be guided through the experiments.
7. Some students did not accept responsibility for their learning.
8. Some students did not like to write the answers.
9. Some students grew to dislike the program toward the end.
10. Some students were bothered by their test results.
11. Some students skimmed through the materials without thinking.
12. Students in slow groups could not concentrate.
13. Some students were hesitant to ask the teacher questions.
14. Some students at their seats watched other students doing experiments.
15. Some students slowed down too much after teacher admonitions.

Teachers:

1. I don't think the teacher is overly fond of my being in class.
2. I gather this teacher resents not having sufficient control over all situations at all times.
3. This teacher does not seem to read the material in advance and plan ahead for the experiments.
4. This was an accelerated group and I had the definite feeling the teacher would have preferred to teach them herself.
5. The teacher felt the program did not provide for individual differences and individual study.
6. The teacher talked too much -- kept interrupting the children.
7. The teacher did not like to grade the tests.
8. The teacher thought her students too dumb to use the programed texts.
9. The teacher's mood varied widely. Some students did not want to ask her questions, probably because of fear, loss of fruit break, etc.
10. The teacher was very insecure initially.
11. Teachers were not always sure students were really learning - particularly because test scores on a percentage basis were low.
12. Teacher relied too heavily on my assistantce -- wanting me to take over the science class.
13. Teachers varied widely in their ability to work with students -- some gave too much freedom -- some too little.

General Methodology:

1. There are so many varied activities going on at one time that it almost requires more time for individual students than one teacher can give.
2. After several months, you have to have numerous experiments set up around the room. This requires space.

Tests:

1. High ceiling of tests was discouraging to students in lower ability groups.
2. The tests were too dependent on vocabulary.
3. For some students, the tests were not accurate (lower ability).
4. The tests varied widely in their quality.
5. The purpose of the sub-tests was not always clear.

B. Positive Comments as Seen by Research Assistants

Materials:

1. Programed materials do obviously have value to increase knowledge of a discipline and to inculcate positive attitudes.
2. Teachers obviously must be trained to use this material.
3. Children also have to learn to use the materials.
4. The novelty of the method and the materials soon wears off, but the teacher can provide the "Hawthorn effect" by providing students with friendly help, by checking on the student's progress from time to time in a way that indicates personal interest, by discussing some concept or experiment with the student, and perhaps by many other ways.
5. Students' progress soon allows you to rank them as to rate of covering the material. This rate will be affected by such variables as level of vocabulary, interest in the material, previous success, the teacher, time of day, the group of which they are members, peer attitudes, goals, values, and many others.
6. It might be a good idea if classes could be combined while using the programed science so as to pool the teacher force...in one room to work under the supervision of the three teachers.
7. Mrs. X feels the programed materials have been of great benefit to her class and has expressed a desire to continue using them after the experiment.
8. With the programed materials, the child is held to a check of whether he understands more frequently than with conventional materials.

9. Mrs. ________ apparently feels the material is suitable for the enrichment of her abler students and that the experiments lead to understanding.
10. While Mrs. ________ indicated a desire to work with the material after the experiment "no strings attached" she seems to have made only slight changes -- occasional class discussion and demonstration of a few experiments. Her enthusiasm for the material is higher than ever -- possibly because she feels she is now important in the process.

Experiments:

1. The best liked feature of the programed science is the experiments.
2. The teacher thought the experiments were very beneficial and taught the children how to work and think independently.
3. The class as a whole said they liked the science period and especially the experiments.
4. An experiment incomprehensible to one child was easily done by another.
5. The students initially had mixed feelings about the experiments. They were more perplexed than anything, yet enthusiastic and cooperative.
6. All in all I would say the children enjoyed the experiments and participated in a number of science experiences which will go a long way towards improving their growth and understanding in science.
7. Mrs. ________ is allowing her better students to continue using the programed text after the experiment. The individual experiments are done by the students along with the regular text.

Students:

1. The interest of the students seemed to be relatively high throughout the experiment in all three classes, even though each of the three classes was conducted in a distinctly different manner.
2. Not too long after starting to use the programed materials, students seemed to find their own working rate, but this was not always the same as their apparent learning rate.
3. The class as a whole said they liked the science period and especially the experiments.
4. To me, one of the most important aspects of the experiment was that students seemd to find their own rate of learning.
5. By the end of the first six weeks, the students had become quite "science minded" as they not only performed the experiments but began to vary them of their own accord to see what would happen.

6. At the end of the experiment, the students still had a strong positive attitude toward the program.
7. The approach encourages outside of school interests in science.
8. Children are perceived to develop a sense of responsibility for their own learning.
9. Students were highly motivated throughout.
10. They loved the experiments.

Teachers:

1. Even though programed material is supposed to provide its own incentive for learning in the form of immediate feedback, I believe that the teacher could have an effect on interest and achievement.
2. It is apparent that Mrs. X feels the programed materials have been of great benefit to her class and she has expressed the desire to continue using the materials after the experiment ends.
3. At first, Mrs. ________ seemed to be rather dependent on the research assistant, but gradually she came to require little if any help.
4. The teacher became increasingly secure.
5. The materials, particularly the experiments, obviously had a very positive in-service education effect on the teachers. They became quite comfortable with the experiments.
6. Excellent rapport was evident between teacher and pupils.
7. The approach is inspirational and instructive for the teacher.
8. After reviewing the total program, she felt that the things learned by her classes outnumbered the criticisms she had.
9. She repeated her favorable attitude toward this type of instruction by continuing through the second semester with them.

General Methodology:

1. Although programed material is no educational cure all it is good medicine to have around.
2. A major benefit of its use is the increasing confidence teachers gain in working with experiments.
3. The teachers themselves learn a great deal of science content in using this medium.
4. Teachers perceive it is possible to have children doing many different things at the same time without chaos resulting.
5. For dealing with individual differences, particularly children achieving above or below their grade level, this approach has considerable potential.

CHAPTER IV
CURRICULUM ANALYSIS

The purposes of the curriculum analysis were (1) to assist in the revision of the initial presentation, (2) to validate the branches, and (3) to discriminate between concept revision for ability and growth levels.

Although some revision had been accomplished as a result of the field testing in 1962-63, data collection specifically for curriculum analysis was begun after the spring experimentation, 1963. The data collection was continued after the fall experiment of the same year. Since a significant relationship exists between the error count and the unit test achievement, a tabulation by frames of the percentage of errors was used to guide in the revision of the initial presentation of material. The second error count only used those frames from programed texts with 1/3 answers given. The percentage of correct responses ranged from a high of 81% for the Science and Its Ways unit to a low of 65% for the Light unit. (See Appendix D for a list of the units and total percentage of correct responses to frames in which no responses were given in the programed text.)

The branches were validated by obtaining a relationship between the level of ability of the student and the branch or route selected. Chi square analyses indicated a significant relationship between the branch and the ability of the student; i.e., in most cases the students of above average ability selected the appropriate route, and the converse situation. However, since the general achievement was not significantly greater for the students using the branching versions, the validation of the branches probably should only apply to the subtests or decision points used in sequencing the students' behavior. (See Appendix E for a listing of the branches and related chi square values.)

The third objective attempted to discriminate between those concepts which are appropriate for certain ability levels and those concepts or facts which are more of a function of growth. A content analysis of the Sound and Light units is described below to indicate any trends which might be found in these materials and those of a related nature. It was hypothesized that: (1) items with low percent of incorrect response and no correlation with I.Q. were satisfactory, or suggested only minor revision on initial presentation, (2) items with low percent of incorrect response and significant correlation with I.Q. suggested remedial branching for children of lower ability, (3) the items with high percent of incorrect response and no correlation with I.Q. suggested the removal of the concept for this grade level, or major revision in the initial presentation, and (4) items with high percent of incorrect response and significant correlation with I.Q. suggested revision of initial presentation and remedial branches for students of lower ability levels.

The percentage of incorrect responses to a group of frames which related to a concept were tabulated. The concept was then classified at a high, medium, or low level of percentage of incorrect responses. The mean I.Q. of 16 classes was then correlated with the percentage of incorrect responses for each concept, using the rank-difference correlation. The correlation between the mean I.Q. and the percentage of incorrect responses is defined as the relationship between the student performance on the concept and the ability level.

TABLE XXVI presents the analysis of selected frames on the Sound and Light units. The relationships between percentage of incorrect responses and ability level, the degree or percentage of frames missed, and the validity of the test item measuring this concept are included in the table.

TABLE XXVI indicates, for example, that the item frames 237-250 which requires a knowledge of more than one relationship (comparing the speeds of sound), is frequently missed by students and does not correlate with ability level. This may suggest that items requiring a knowledge of such abstract relationships is not appropriate at the fourth grade level. However, an item requiring a knowledge of only one relationship (frames 275-282, light travels faster than sound) is missed much less frequently and has a significant relationship with ability. In these cases, it may be more beneficial to give remedial branches to only those of average and lower ability levels.

TABLE XXVI

ANALYSIS OF SELECTED FRAMES ON THE SOUND AND LIGHT UNITS

Frames	r	information	level of % incorrect	test item measuring information in frames bis$_r$	 diff index
		SOUND			
11-18	.02	sound waves	medium	.68	.63
23-33	.62*	objects vibrate	low	.37	.49
112-125	.67*	pitch is the highness and lowness of sounds	medium	.86	.55
237-250	-.32	comparing speeds of sound in gases, liquids & solids	high	.55	.86
253-264	.58*	" " "	medium	.30	.96
275-282	.77*	light travels faster than sound	medium	.68	.75
295-301	.58*	echoes	low	.53 .70	.74 .72
		LIGHT			
5-19	.65*	man-made and natural light	low	.60 .40	.52 .52
95-104	.50	filament inside light glows when electricity passes through it and produces light	high	.73 .63	.57 .46
108-117	.65*	filaments are coiled to produce more light	high	.73 .63	.57 .46
265-279	.45	refraction is the bending of light	low	.61 .79	.81 .63
351-361	.63*	color is due to the reflection of certain light rays	medium	.80	.62
408-416	.62*	basic colors; lens refract light	medium	.60 .86 .81	.57 .47 .53

*significant at .05 level.

In summary, the content analysis of the Sound and Light units revealed that nine groups of items had a significant relationship with ability levels and the percentage of incorrect pupil responses. In such cases, branching for individualization is indicated. A high or low level of percentage of incorrect responses does not necessarily imply a significant relationship with ability. Items which had a high level of percentage of incorrect responses and no significant relationship with ability are probably not appropriate for this grade level. Items with a low level of percentage of incorrect response and no significant relationship with ability are most likely satisfactorily presented in the units.

An additional analysis of the programed science materials involved a content analysis of the Sound unit in an attempt to show if there were any similar characteristics in the unit on Sound of these materials which contribute to correct and incorrect recall by the entire sample on items of the unit test.

The Sound unit test was analyzed for items correctly or incorrectly recalled by the entire sample. A content analysis of the unit on Sound was the means by which the nature of the programing revealed any commonalities which differentiate the content of the materials relating to the items correctly recalled from the items incorrectly recalled by the entire sample.

A content analysis of the Sound unit test showed that no individual test items were recalled correctly or incorrectly by the entire sample. Since there were no single test items correctly or incorrectly recalled by the entire sample, this investigator analyzed those items on the Sound unit test most frequently recalled correctly by the sample; i.e., the test items recalled correctly by at least 94 per cent of the sample, and those test items recalled correctly by only a small portion of the sample; i.e., 32 to 49 per cent of the sample. This analysis determined any common characteristics in the nature of programing the Sound unit which were associated with those test items recalled correctly and incorrectly by most of the sample.

TABLE XXVII on page 93 shows the items recalled correctly and incorrectly by the majority of the sample. A content analysis of the programed science materials, as shown in TABLE XXVII, shows that the only characteristic associated with the items correctly recalled by at least 94 per cent of the sample is that the nature of the item is a low level of cognition; i.e., $S_1 - R_1$ relationship or basic one to one association of a stimulus to a response. The only characteristic of the programed materials associated with at least 32 per cent of correct recall by the sample was related to a high level of cognition or abstractness; i.e., generalizations and discrimination of responses and stimuli. Correct recall of an item, therefore, hinges on the nature of the item itself; i.e., cognitive level of the information should be at a low level in order for the majority of the pupils to recall successfully with programed materials, and the converse is true if the item is at a high level of cognition or abstractness. Abstract information, however, could be programed successfully if it were presented in a concrete or understandable fashion and information presented at a high abstract level were lowered to a very simple level of cognitive presentation.

Initial revision and field testing indicate that programed materials are readily adaptable to an analysis which will assist in the efficient presentation of content to the student. However, the numerous steps in the crude stages of revision, in addition to the number of significant programing variables which must eventually be considered, makes this process one of a time-consuming but essential nature.

TABLE XXVII

CONTENT ANALYSIS OF ITEMS MOST FREQUENTLY RECALLED CORRECTLY AND INCORRECTLY ON SOUND UNIT TEST

Nature of programed science materials	1	2	3	4	5	6	7	8	9
Number missed	4	4	4	5	5	46	48	55	51
Number correct	83	83	83	82	82	41	39	26	36
Per cent correct	95	95	95	94	94	47	45	32	41
Level of cognition	lo	lo	lo	lo	lo	hi	hi	hi	hi
Nature of stating item on test	cp	cp	ft	ft	ft	cp	cp	ft	ft
Number of different facts contributing to recall of item	2	0	1	1	2	0	0	2	2
Number of different concepts contributing to recall of item	6	6	0	1	0	4	4	1	0
Total number of facts contributing to recall of item	13	0	5	10	26	0	0	13	22
Total number of concepts contributing to recall of item	40	28	0	4	0	29	29	10	0
Total number of facts and concepts contributing to recall of item	53	28	5	14	26	29	29	33	22

Where: lo means low level of cognition
hi means high level of cognition
cp means concept
ft means fact

CHAPTER V
PROGRAMED MATERIALS WITH ELEMENTS OF CRITICAL THINKING

I. STATEMENT OF THE PROBLEM

An amendment to Project #1972 granted a one-year extension to investigate the influence of techniques of critical thinking which could be introduced into the programed materials. Development and previous experience with the programed materials, coupled with the laboratory activities, suggested their use in the investigation of the critical thinking variable in the following ways:

1. the rather radical approach to teaching elementary science is administratively feasible within the context of a real classroom,
2. the materials offer control of the content and its mode of presentation,
3. the laboratory experiences provide a means of deliberately moving from the concrete to the abstract.

Suchman (27) recognizes the problem of control of teacher variation in his training program on the structure and function of inquiry. Modification of the programed units in which techniques of critical thinking are included allows the self-instructional process which is possible using the programed units without the critical thinking revisions. Thus, by assigning some students in each of the experimental and comparison groups to the same classes, it seemed likely that most of the systematic effects would operate in the same way on both groups. In addition, the self-instructional process allowed more control of scientific content and the methods of presenting the techniques of critical thinking.

The laboratory experiences were considered crucial in assisting the student to abstract critical thinking operations. The findings of Inhelder and Piaget (16) stress moving from the concrete to the abstract in building formal logical operations. In the development of scientific inquiry Suchman (27) emphasizes the necessity of performing concrete operations to provide an empirical ordering of objects and events. Although the need to move from concrete activities in order to develop some capacity for scientific thinking is widely recognized, further study of the growth process is suggested by the findings and recommendations of the researchers engaged in a study of this problem.

Objectives. Two primary objectives were set forth for this pilot study investigating the influence of techniques of critical thinking:

1. To determine the relative effects on critical thinking ability between a group of fourth-grade students who have studied programed materials modified to include elements of critical thinking with a group of fourth-grade students who studied the programed materials without elements of critical thinking deliberately and specifically included.

2. To determine the relationship among the variables, critical thinking, intelligence, personality, and science achievement.

The techniques of critical thinking were patterned after the work of Ennis[1], who defines the basic notion of critical thinking as "the correct assessing of statements." Ennis lists and discusses 12 aspects of critical thinking which describe ways to avoid pitfalls in assessment. Those characteristics of critical thinking applicable to Suchman's stages of inquiry: searching, data processing, discovery and verification, are included in the programed materials. Presented below is a brief description of Suchman's stages in inquiry and the characteristics of critical thinking which were identified with each stage.

Suchman describes the act of searching as the intake of data planned and controlled by the observer. Two characteristics of searching are mobility, or the exposing of oneself to a range of data sources, and manipulation, or increasing the yield from a source of data. He also states that the searching organism must have a domain in which to operate, freedom to move within this domain, a "set" to guide the selective procedures, and an overall system or plan[2]. The critical thinking elements included in the modified programed versions did not, however, allow the flexibility that was encouraged in the searching action in Suchman's inquiry program. The training was primarily directed toward the formulation of the problem. This included three of Ennis' characteristics: judging whether (1) the problem has been identified, (2) something is an assumption, and (3) a definition is adequate.[3]

---

[1] Robert H. Ennis, "A Concept of Critical Thinking," Harvard Education Review, XXXII (Winter, 1962), p. 83.

[2] Richard Suchman, The Elementary School Training Program in Scientific Inquiry, College of Education, University of Illinois, Urbana, Illinois, April, 1963, p. 7.

Suchman's second stage of inquiry, data processing, is described as the ability to assimilate data within the framework of conceptual patterns. Four types of data processing are: analysis, comparison, isolation, and repetition. The aspects of critical thinking included from Ennis in this stage of inquiry are: judging whether (1) there is ambiguity in a line of reasoning, (2) certain statements contradict each other, and (3) an observation statement is reliable.[4]

Discovery, which ends in a higher degree of assimilation than was possible prior to the act, and verification, are defined by Suchman as the last two stages of inquiry. The following judgments listed by Ennis are included in these stages: judging whether (1) a conclusion follows necessarily, (2) a statement is actually the application of a certain principle, and (3) an inductive conclusion is warranted.

---

[3] Ennis, op. cit., p. 12.

[4] Ibid.

II. METHOD AND PROCEDURE

Materials Development

During the summer and fall of 1964, the programed units of Sound, Light, Heat, and Weather, were revised to include selected elements of critical thinking. Programing methods remained in the revised units, although the presentation was generally less structured.

Specifically, the first unit, Sound, introduced the student to asking a question, and thinking about the evidence and/or relating the question to an event before supplying an answer. He was first presented with a simple situation, a question, and choice of response (yes, no, or maybe). An attempt was made to develop an awareness in the student of the need for the evaluation of evidence before answering a question. The laboratory experiments introduced the notion of a question as a hypothesis, and the student observed and gathered data to assist him in drawing a conclusion. Questions concerned with what he observed during the laboratory experiment related to whether he had identified the problem and why he arrived at his conclusion. The process of answering "why" to a question was emphasized in this unit, rather than acquisition of content.*

---

*See Appendix F for sample of material.

The second unit, Light, extended the student's experiences to making guesses (formulating hypotheses) and formalized some data processing techniques. Practice material in classification of facts and data into sets and subsets was incorporated to encourage formulation of a more intelligent guess (hypothesis) about the solution of a problem. Questions about the relationships of sets and subsets provided practice in logical operations, which introduced the students to the ideas of generalization and causality. The concept of an assumption was introduced into the process of stating a hypothesis and arriving at a tentative conclusion. An awareness of the relationships between events or groups was used as a basis to show the relationship between the hypothesis, assumption, and conclusion. The scientific content presented in this unit provided background for making an hypothesis or stating an assumption. The experiments emphasized the detection of assumptions, the demonstration of evidence, and the application of principles. The logical operations served as a basis for the deductive operations.

The Heat unit utilized the experiments as sources for gathering data, discussing causality, and generalizing conclusions. Gathering evidence and generalizing findings from several activities emphasized the use of inductive processes.

The last unit, Weather, was written to eliminate most of the critical thinking props. The student was left relatively independent in the processes of formulating hypotheses, observing data, and drawing conclusions. The opportunity to apply some principles from the Heat unit was provided. In other words, the student was expected to apply some of the facets of the process of critical thinking in the solution of problems and in the application of scientific principles to the solution of these problems.

Sample. One hundred sixty subjects were randomly selected from a population of 240 fourth-grade students of at least average ability. One hundred sixty students in nine classes were randomly assigned to either an experimental or control group. The other eighty students were given conventional programed materials but, unknown to them, were not to be included in the data analysis.

The experimental group studied modified programed science units which included the elements of critical thinking as described previously. The control group studied the programed units without the specific inclusion of techniques of critical thinking. The experimental period, of two months duration, was initiated in December, 1964.

Data Collection. The following tests were administered to all subjects: (1) a critical thinking test, (2) a true-false science achievement test, (3) the STEP Science Achievement test, and (4) the California Test of Personality, 1953 Revised Form AA. The Lorge-Thorndike intelligence scores were secured from the subjects' cumulative records, since these data had been gathered earlier in the year. The critical thinking test (58 logistic and inferential items) and the true-false science achievement tests were prepared and validated by the staff of the Bureau of Educational Research, University of Virginia. (See Appendix G).

III. ANALYSIS OF DATA

Initial Measures. To determine if any significant mean differences existed between the experimental and control groups on the initial test measures, critical thinking, intelligence, and science achievement, t-tests were used. TABLE XXVIII summarizes the analysis of pre-experimental testing. The means, standard errors and critical ratios are given for each of the four initial measures.

TABLE XXVIII indicates that no initial significant mean differences in critical thinking, intelligence, and science achievement were found at the .05 level of significance between the experimental and control groups.

TABLE XXVIII

SIGNIFICANCE OF DIFFERENCE BETWEEN THE MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS ON FOUR INITIAL MEASURES

Critical Thinking

Experimental
$M = 42.2$

Control
$M = 42.1$

Mean difference = .1
$SE_D = .71$
$CR = .14$

Accept null hypothesis

Intelligence

Experimental
$M = 113.2$

Control
$M = 111.3$

Mean difference = 1.9
$SE_D = 1.78$
$CR = 1.07$

Accept null hypothesis

True-False Science Achievement

Experimental
$M = 128.2$

Control
$M = 127.2$

Mean difference = 1.0
$SE_D = 9.45$
$CR = .11$

Accept null hypothesis

STEP A Science Achievement

Experimental
$M = 41.1$

Control
$M = 40.3$

Mean difference = .8
$SE_D = 1.30$
$CR = .62$

Accept null hypothesis

Critical Thinking Ability

A t-test was used to compare the means of the post-experimental or final measure of critical thinking of the experimental and control groups. TABLE XXIX summarizes the mean comparison between the two groups.

TABLE XXIX

A MEAN COMPARISON BETWEEN THE EXPERIMENTAL AND CONTROL GROUPS ON THE FINAL MEASURES OF CRITICAL THINKING ABILITY

Final Testing Experimental Group	Final Testing Control Group
N = 80	N = 80
M = 45.2	M = 43.5
SD= 3.49	SD= 5.65

Mean difference = 1.70

SE_D = .74

CR = 2.30

Reject Null Hypothesis

.05 = 1.96

.01 = 2.58

Conclusion: Students who studied programed science materials which included elements of critical thinking improved significantly, in their ability to think critically, over those students who studied the same programed science materials from which the elements of critical thinking were eliminated.

TABLE XXIX indicates that the null hypothesis of no mean difference in critical thinking ability was rejected at the .05 level of significance. The data suggest, therefore, that positive benefit was derived from the inclusion of elements of critical thinking in the science materials.

In order to determine any significant mean gains, correlated t-tests were run between the pre-experimental and post-experimental means of critical thinking ability for the total sample, the experimental group, and the control group. TABLE XXX presents an analysis of the pre- and post-experimental means of critical thinking ability.

TABLE XXX

MEAN COMPARISON BETWEEN PRE- AND POST-EXPERIMENTAL MEANS OF CRITICAL THINKING ABILITY FOR THE TOTAL SAMPLE, THE EXPERIMENTAL GROUP, AND THE CONTROL GROUP

Pre-experimental		**Post-experimental**
	Total Sample	
N = 160 M = 42.1 SD = 4.50	Mean difference = 2.20 SE_D = .72 CR = 6.11	N = 160 M = 44.3 SD = 4.77 **Reject** null hypothesis
	Experimental Group	
N = 80 M = 42.2 SD = 3.86	Mean difference = 3.00 SE_D = .39 CR = 7.69	N = 80 M = 45.2 SD = 3.49 **Reject** null hypothesis
	Control Group	
N = 80 M = 42.1 SD = 5.07	Mean difference = 1.40 SE_D = .61 CR = 2.30	N = 80 M = 43.5 SD = 5.65 **Reject** null hypothesis

The data presented in TABLE XXX indicate that the mean gain in critical thinking after the two-month experimental period was found to be significant at the .01 level for the total sample and the experimental group, and significant at the .05 level for the control group. Both the experimental and control groups showed evidence of significant mean gains in critical thinking, although the experimental group evidenced a greater mean gain.

The experimental and control groups were each divided into two subsets in order to present more specifically the nature of the mean gain in critical thinking demonstrated by the experimental and control groups. TABLE XXXI presents intelligence and critical thinking data for the four groups. Groups I and IV compose the experimental group, or all students in these two groups have used the critical thinking materials. However, the mean intelligence is higher in Group I than Group IV. The control group is similarly divided into Groups II and III.

TABLE XXXI

EXPERIMENTAL AND CONTROL GROUPS EACH DIVIDED INTO SUBSETS OF HIGHER AND LOWER ABILITY LEVELS WITH RELATED I.Q. AND CRITICAL THINKING DATA

		Intelligence		Critical Thinking	
Group	N	M	SD	Mean Gain	CR
I	48	119.2	8.37	3.00	5.77**
II	48	119.1	8.48	1.60	2.96**
III	32	99.7	7.07	1.31	1.02
IV	32	104.1	3.18	2.80	4.83**

Groups I and IV = Experimental Group
Groups II and III = Control Group

**Significant at .01 level

It can be observed from TABLE XXXI that Groups I, II, and IV demonstrated a significant mean gain in critical thinking ability at the .01 level of significance. The null hypothesis of no mean difference between pre-experimental and post-experimental critical thinking measures was accepted for Group III. The higher ability group using the critical thinking materials (Group I) experienced a greater mean gain than any of the other three groups. The high ability using the programed materials without elements of critical thinking (Group II) did, however, make a lesser but significant mean gain in critical thinking ability; but Group III, composed of students with somewhat lower intelligence scores who were in the control group were not able to demonstrate any mean gain. In a comparison of the two lower ability groups, one might expect Group IV to demonstrate a greater mean gain than Group III because of a higher mean intelligence, although the positive influence of the critical thinking units on Group IV is indicated by a greater mean gain in critical thinking ability than Group II.

Relationship of Critical Thinking with Psychological Variables

In a related study[5] the students were divided into high and low critical thinking groups. The high critical thinking group was composed of 27 percent of the students (N = 43) who scored the highest on the critical thinking test; the low critical thinking group was composed of 27 percent of the students (N = 43) who scored the lowest on the critical thinking test. The high and low critical thinking groups were compared on measures of critical thinking, intelligence, and personality (personal adjustment, social adjustment, and total adjustment, as measured by the California Personality Test). The high critical thinking group performed significantly higher on each of the variables than did the low critical thinking group.

---

[5] Panayiotes D. Constantinides, "The Relationship of Critical Thinking Ability to Intelligence, to Personality, and to Teacher Evaluation of Pupil Personality," unpublished Doctoral Dissertation, University of Virginia, Charlottesville, Virginia, 1965, p. 61 ff.

The intelligence scores of the high and low critical thinking groups were correlated with the critical thinking scores (r = .50), and a critical thinking score was predicted for each student from his intelligence measure. A comparison of the student's actual critical thinking score with his predicted score determined his classification as an overachiever or underachiever in critical thinking; e.g., actual score greater than predicted or expected score would designate the student as an overachiever. A comparison of the performance of just the overachievers in the high and low critical thinking groups again indicated significantly better scores for the high group in critical thinking, personal adjustment, social adjustment, total adjustment, and intelligence.

A comparison of the underachievers in the high and low critical thinking groups indicated a significant difference in only two variables, critical thinking and intelligence. It appears that personality does not hold a simple relationship with high and low critical thinking ability but varies with the level of critical thinking ability relative to his expected performance.

The relationships among critical thinking, intelligence, personality, and science achievement are summarized in TABLE XXXII below.

TABLE XXXII

THE INTERRELATIONSHIPS AMONG THE CRITICAL THINKING, INTELLIGENCE, PERSONALITY AND SCIENCE ACHIEVEMENT VARIABLES, AND THE MEANS AND STANDARD DEVIATIONS OF EACH VARIABLE

Variables:
1. Critical Thinking
2. Intelligence
3. Personality Adjustment
4. True-False Science Achievement

Zero-order Correlations

		CT 1	IQ 2	PA 3	TF 4
CT	1	1.00	.50**	.29**	.37**
IQ	2		1.00	.39**	.18*
PA	3			1.00	.23**
TF	4				1.00
		(1)	(2)	(3)	(4)
M =		42.1	112.2	110.1	127.7
SD =		4.50	11.31	18.13	18.90

158 df = .05 = .16
.01 = .21

The zero-order correlation coefficients were used to calculate partial Beta weights. These values have been substituted in the equation below in order to determine the percentage contribution of the independent variables (2 = I.Q., 3 = personality, 4 = achievement) to the variance of critical thinking (1 = critical thinking)

$$R^2_{1(234)} = r_{12}\,B_{12.34} + r_{13}\,B_{13,24} + r_{14}\,B_{14.23}$$

$$= (.50)(.41) \quad (.29)(.062) \quad (.37)(.316)$$

$$= .206 + .018 + .117$$

$$= .341$$

As shown above $R^2_{1(234)} = .34$; or 34% of the variance in critical thinking can be attributed to differences in general intelligence, personality adjustment, and achievement. Most of the variance of critical thinking can be accounted for by the intelligence variable (about 21%), while the contribution of achievement is about 12%, and personality adjustment contributes less than 2%.

The r = .37 between the True-False achievement and the critical thinking test may suggest the inclusion of achievement items which require a higher level of mental process; however, a higher relationship of achievement with critical thinking would not necessarily increase the value of the multiple correlation coefficient, since an achievement test which requires more thought would likely intercorrelate substantially with the intelligence variable. The small independent contribution of the personality variable to critical thinking can be accounted for by the relatively higher relationship of personality with intelligence than with critical thinking, the relationship of the personality variable with high critical thinking and with critical thinking ability relative to intelligence. Thus, the influence of the personality variable might be expected to vary with intellectual and critical thinking abilities; e.g., account for a relatively greater percentage of the critical thinking variance when subjects are of high critical ability than when of low ability. The remaining 66% of the variance of critical thinking must be attributed to factors not measured in the regression equation above.

IV. FINDINGS AND CONCLUSIONS

1. Fourth-grade students of above average ability had a significantly greater mean performance in critical thinking ability after two months study of elements of critical thinking introduced into four units of programed science materials than did fourth-grade students of above average ability who studied the programed units without the specific inclusion of critical thinking techniques.

2. Students with intelligence scores falling approximately between average and + 1 SD did not make a significant

mean gain in critical thinking ability unless they had studied the critical thinking materials. Students falling above + 1 SD in intelligence demonstrated a significant mean gain in critical thinking ability, irrespective of the materials, although the experimental group using these materials demonstrated a greater mean gain.

3. Students with high critical thinking ability evidence significantly higher intelligence, personal adjustment and social adjustment than do students of low critical thinking ability.

A comparison of students (with a relatively greater critical thinking score than is expected from intelligence) in high and low critical thinking groups indicates superior adjustment for these students in the high critical thinking group. A comparison of students whose critical thinking scores are less than expected from intelligence, and who fall in high and low critical thinking groups does not indicate any significant differences in personal and social adjustment.

4. An $R^2_{1(234)} = .34$ in a multiple correlation, in which critical thinking is defined as the criterion variable, and intelligence, personal adjustment, and science achievement are defined as the dependent variables. About 21% of the variance is attributed to intelligence, and about 12% to achievement. Approximately 66% of the variance of critical thinking must be attributed to factors not measured in the above multiple correlation.

CHAPTER VI
CONCLUSIONS AND IMPLICATIONS

1. Initial experimentation and field testing by the investigators suggested that programed science materials, coupled with the individual performance by students of laboratory experiences, had promise as one instructional method in the modern elementary school curriculum. This conclusion was based on (1) the science achievement data which either favored the experimental group or was comparable to the more traditional approach, (2) the demonstrated ability of children to conduct their own experiments and the administrative feasibility of such an approach, and (3) the generally favorable attitudes of students and teachers, even after an experimental period of one year.

Both students and teachers had similar expressions of approval for the method of presentation. The experiments, the individual pacing, and small-step sequencing with immediate reinforcement elicited favorable responses by students and teachers. The teachers felt the materials had an instructional benefit for them. Not only had they learned some science themselves, but no longer feared laboratory activities. Perhaps their more favorable attitude toward science was partly responsible for the attitude frequently expressed by students that they enjoyed learning science. Expressions of dislike had similar reactions among most students and teachers, and arose when the method of presentation was too difficult for the below average reader. About ten percent of students and teachers reacted very negatively to programed instruction.

2. A more formal evaluation of the materials over a period of four and one half months compared three groups of fourth-grade students: (1) programed materials with individual performance of laboratory experiences, (2) programed materials with students reading about laboratory activities, and (3) a textbook approach with teacher demonstrations of laboratory activities. The three approaches were evaluated by the measures of instructional effectiveness discussed below:

Achievement and Retention. No significant mean difference in achievement (as measured by the STEP Science Achievement Test and a true-false test constructed by the Research Bureau) in post-experimental testing and after a period of four and one half months. Although the programed materials did not evidence superiority over the more traditional approach, the reader should note that all groups were compared under a favorable instructional situation; i.e., assistance in materials collection, instruction, or grading were provided to all teachers.

Laboratory Resourcefulness. To determine any direct influence of the student's performing the laboratory activities, the Research Bureau personnel constructed and administered a laboratory resourcefulness test. A significantly greater mean performance was found for the programed instruction group which individually performed the laboratory experiences. A significant interaction effect was found between intelligence levels and treatments, in which the brighter children who individually performed the experiments, evidenced a proportionally greater mean achievement than children of the same ability level in either of the other two groups.

Interest and Level of Interest. No significant mean differences or interaction effects with ability were found among the three groups.

Attitude. Expressions of students' attitudes were gathered at periodic intervals throughout the experimental period. The attitudes expressed by the group using programed materials and only reading about the experiments, were significantly lower than the other two groups. Strangely, this less than enthusiastic attitude did not appear to influence the students' achievement, since the mean achievement did not differ significantly among the three groups. Could it be that attitudes or motivation do not have the same effect on pupil performance of programed materials as opposed to the more traditional method of instruction?

Attitudes remained enthusiastic for both the programed-with-laboratory-activities and the more traditional groups during the experimental period. Certainly for the programed group, the laboratory experiences appeared to have motivational value. A relationship existed between rate of progress of the students and the nature of their attitudes. Students who were making little progress had relatively lower achievement and generally expressed more dislike for the program. These were usually children of lower ability who found the vocabulary or reading level too difficult. Children with similar but somewhat higher ability demonstrated about the same science achievement, but expressed more favorable attitudes. Quite obviously, motivation factors inherent in programed materials are less effective if the student is not learning. One exception was an extremely low ability class in which, although the students were obviously not understanding, they had positive attitudes. The best we could conjecture was that the relative freedom which the students had during "science" more than counteracted the meaninglessness of the programed materials. The evidence indicates that the teacher does not affect attitudes in the same degree with the programed learning situation as she is able to under the more traditional approach.

3. The influence of the response factors; (1) constructed response vs. multiple choice, (2) all answers vs. partial feedback, and (3) linear vs. branching, were evaluated using the measures of instructional effectiveness described above. No significant mean differences or interaction effects were found among these programed versions.

A related study obtained the relationships among the variables, time, achievement, reading level, and intelligence. Findings suggest that student performance differs when using the different versions. After holding intelligence and time constant, significant positive relationship was found to exist between achievement and reading level for students using the programed versions with partial feedback, but not for the students using the all-answer versions. In another instance, a significant negative relationship existed between time and achievement, with the effects of reading and intelligence held constant, but only for those students using the constructed response versions. No mean differences in achievement, reading level, or intelligence were found among the groups in the study; however, the linear version was completed in significantly less time than the branching texts.

The differences in student performance support Lumsdaine's (18) view that intermediate-level principles or interactions among variables need to be determined. Perhaps the difference in the relationships between achievement and reading level in the one-third and all-answer versions can be attributed to an element of prompting allowed by the all-answer version. Differences in performance between the constructed response and multiple choice versions may reflect Pressey's (21) notion that multiple choice items in a programed text are not of the same nature as a test item; e.g., the items are elucidative or corrective.

The additional research to develop intermediate principles must not only include programing variables, but must also consider variables related to the nature of the content and the learner. The curriculum analysis first considered student performance of frames and related this performance with ability levels. Although a significant relationship exists between percentage of correct responses and achievement, the correct responses may or may not be related to the ability of the student. This relationship will vary with the nature of the concept, the reading level or vocabulary involved, and the type of presentation.

Further refinements will doubtless define stimulus and response characteristics which are most appropriate for a given concept type and student characteristic. Small step, logically organized presentations are successful with knowledge of results or confirmation, while a less ordered or larger step approach may be compensated for by an explanation of a correct or incorrect response. Likewise, step size can be increased without loss in criterion achievement if responses are meaningful. There is evidence that overt responses may interfere with complete prompting, while prompting or cuing within limits appears successful, and overt responses seem more successful with confirmation. Continued investigation of these programing variables for various curriculums and students may eventually produce principles of programing and a theory of instruction (18) or lead to more effective methods of translating theories of learning into curriculum development (24).

4. Critical thinking. A significantly greater mean achievement in critical thinking ability was evidenced by those students who had studied the materials with elements of critical thinking. The materials with the laboratory experiences seem to be adaptable in assisting students to move from concrete to abstract logical and inferential operations. Students with intelligence scores ranging between average and + 1 SD seem to need specific training in critical thinking before any significant mean gain in these skills can be realized. A multiple correlation analysis reveals that intelligence is the largest component of critical thinking; however, personality appears to be a significant variable for the high critical thinkers.

SUMMARY

The investigators conclude, after almost three years of work with programed science experiments coupled with student performed experiments, that this approach can be a valuable adjunct to a modern elementary school science program. This is particularly true in the sense that students are apparently very highly motivated over long periods of time because of the opportunity to proceed at their own rate and in particular to do the many experiments BY THEMSELVES.

The investigators were not able to demonstrate that this approach is superior to conventional teaching (as defined in the study). They feel this, in part, was due to weaknesses in their programing techniques rather than in the method per se.

Valuable side effects of this approach to teaching science included: (1) the in-service training benefits for teachers in science, (2) the confidence obviously developed by teachers for doing experiments rather than avoiding them, (3) the aid for those teachers who had not learned how to free individuals for small group and individualized work in the classroom, (4) the evident shouldering of more responsibility by children for their own learning, (5) the strengthening of a science program where teachers are especially untrained in science, (6) the opportunity for dealing more effectively with those children unusually far above or below the class in achievement in science, and (7) the lack of necessity for duplicate sets of science equipment.

The investigators have serious reservations about programed materials constituting the totality of any elementary school science program. It was apparent from those teachers who chose to continue with the materials when the formal experimentation period was over, that there are a variety of ways in which short programed units can become another facet (another instructional tool) for the balanced science program in a modern elementary school.

BIBLIOGRAPHY

1. Bryant, Paul Payne, "Science Understandings Considered Important for Children and the Science Required of Elementary Teachers." Unpublished Doctoral Dissertation, George Peabody College for Teachers, 1959.

2. Carpenter, Regan, "A Study of the Effectiveness of the Problem Solving Method and the Textbook-Discussion Method in Elementary Science Instruction." Unpublished Doctoral Dissertation, University of Colorado, 1958.

3. Carr, W.J., "A Review of the Literature on Certain Aspects of Automated Instruction," appeared in Lumsdaine, A.A. and Glaser, Robert (ed.) *Teaching Machines and Programmed Learning*, Washington, D.C.: NEA, 1960.

4. Constantinides, Panayiotes D., "The Relationship of Critical Thinking Ability to Intelligence, to Personality, and to Teacher Evaluation of Pupil Personality," Unpublished Doctoral Dissertation, University of Virginia, 1965.

5. Coulson, J.E. and H.F. Silberman, "Automated Teaching and Individual Differences," *Audio Visual Communication Review*, 9(1):5-15, January-February, 1961.

6. ______., *Results of Initial Experiment in Automated Teaching*. System Development Corporation, Santa Monica California, July, 1959.

7. Crabtree, Joseph F., II, Independent Study, Division of Educational Research, University of Virginia, December, 1963.

8. Craig, Gerald S., *Certain Techniques Used in Developing a Course of Study in Science for the Horace Mann Elementary School*. New York: Bureau of Publications, Teachers College, Columbia University, 1927.

9. Dubins, Mortimer Ira, *Current Practices in Elementary School Science with Reference to Courses of Study, Published from 1940 to 1952, and the Extent of Activities Undertaken for the Improvement of Instruction*. Doctoral Dissertation, Boston University, 1953.

10. Dutton, Sherman S., "An Experimental Study in the Programming of Science Instruction for the Fourth Grade." Unpublished Doctoral Dissertation, University of Virginia, 1962.

11. Ennis, Robert H., "A Concept of Critical Thinking," *Harvard Education Review*, XXXII, Winter, 1962.

12. Fry, E.B., *Teaching Machines: An Investigation of Constructed Versus Multiple-Choice Response Modes.* Doctoral Dissertation, University of Southern California, 1960.

13. Gilbert, Alice, "Science Content for the Elementary Schools." *School Science and Mathematics*, 43, November, 1943.

14. Hedges, William D. and Mary Ann MacDougall, "An Investigation of the Status of Science Education in Selected Public Elementary Schools of Virginia," Division of Educational Research, University of Virginia, Charlottesville, Virginia, September, 1962.

15. Hughes, J.H., "The Effectiveness of Programmed Instruction: Experimental Findings for 7070 Training," International Business Machines Corporation, New York, January, 1961; reported in *Programmed Instruction*, I, May, 1951.

16. Inhelder, Barbel and Jean Piaget, *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books, Inc., 1958.

17. Keislar, E.R. and John D. McNeil, "Teaching Scientific Theory to First Grade Pupils by Auto-Instructional Device." University of California, Los Angeles, 1961.

18. Lumsdaine, A.A., "Educational Technology, Programed Learning, and Instructional Science," Chapter XVI, *Theories of Learning and Instruction*, The Sixty-third *Yearbook of the National Society for* the Study of Education, Edited by Ernest R. Hilgard, The University of Chicago Press, Chicago, Illinois, 1964.

19. Porter, D., "Some Effects of Year Long Teaching Machine Instruction," in E.H. Galanter (ed.), *Automatic Teaching: The State of the Art*. John Wiley and Sons, Inc., New York, 1959.

20. Pressey, S.L., "A Simple Apparatus Which Gives Tests and Scores -- and Teaches." *School and Society*, 23, March 20, 1926.

21. ______., "Autoinstruction: Perspectives, Problems, Potentials," Chapter XV, Theories of Learning and Instruction, The Sixty-third Yearbook of the National Society for the Study of Education, Edited by Ernest R. Hilgard, The University of Chicago Press, Chicago, Illinois, 1964.

22. ______., "Development and Appraisal of Devices Providing Immediate Automatic Scoring of Objective Tests and Concomitant Self Instruction." Journal of Psychology, 29, April, 1950.

23. ______., "Some Perspectives and Major Problems Regarding Teaching Machines," appeared in Lumsdaine, A.A., and Glaser, Robert (ed.) Teaching Machines and Programed Learning, Washington, D.C.: NEA, 1960.

24. Silberman, H. and J. Coulson (and R. Melaragno and G. Newmark). Use of Exploratory Research and Individual Tutoring Techniques for the Development of Programming Methods and Theory. TM-895/200/00, System Development Corporation, Santa Monica, California, 1964.

25. Skinner, B.F., "The Science of Learning and the Art of Teaching." Harvard Education Review, XXIV, Spring, 1954.

26. ______., "Teaching Machines." Science, 128, October 24, 1958.

27. Suchman, J. Richard, The Elementary School Training Program in Scientific Inquiry. College of Education, University of Illinois, Urbana, Illinois, April, 1963.

28. Taylor, Alton L., "The Readability of Elementary Programed Science Materials," The Field Testing of a One-Year Course of Study Using Programed Science Materials at the Fourth Grade Level, Division of Educational Research Bulletin No. 1, University of Virginia, January, 1964.

APPENDICES

Appendix		Page*
A	Laboratory Resourcefulness Test	1
B	Special Adjustments for Lowest Ability Students	6
C	Initial and Final Means for Eight Versions	9
D	Percent Correct Responses Per Unit with No Responses Given	11
E	Branches Per Unit and Chi-Square Values for Each Branch	13
F	Sample of Sound Unit Critical Thinking	15
G	True - False Test . . . . Critical Thinking Test.	20 30

*This numbering appears only on the cover sheet for each appendix and is not to be confused with the numbering from the original source which appears within the appendices.

APPENDIX A

Laboratory Resourcefulness Test

How Resourceful Are You?

During this school year you have been studying science. You have done some experiments. Below are listed some other experiments for you to do. Can you figure out how to do these experiments? If you can, please do not tell your classmates for they need a chance to see if they can figure them out.

You will not be graded on this. We want to find out if fourth grade children can do experiments like these. Here is the first experiment:

Experiment 1

Given: One wire about 8 inches long.

Problem: To make two wires that are the same length using only your bare hands. Do not use a ruler or anything.

How? ______________________________

Experiment 2

Given: Two bottles that look different, but are about the same size. A pan of water.

Problem: Which bottle holds the most water?

How? ______________________________

Experiment 3

Given: One milk bottle full of water on a table and one empty milk bottle on the floor. Also a piece of rubber hose about three or four feet long.

Problem: To get the water out of the bottle on the table and into the bottle on the floor without moving the bottles.

How? ____________________

Experiment 4

Given: A string one yard long; scissors.

Problem: To get a piece of string 6 inches long without using anything but the scissors and string.

How? ____________________

Experiment 5

Given: Magnifying glass or lens (thick in the middle) and something to read.

Problem: To find out which way light rays bend when the glass is thick in the middle.

How? ____________________

Experiment 6

Given: Water, a measuring cup like you cook with and two small rocks or pebbles. The rocks should look about the same size but the teacher should know that one of them is really a little bigger than the other.

Problem: To find out which of the rocks takes up the most space.

How? ____

____

____

____

Experiment 7

Given: A balance like they use in chemistry class or a homemade balance like a see-saw with a paper cup glued to each end like this: Also, one rock or stone, and some sand.

Problem: To get a pile of sand that weighs only half as much as the rock.

How? ____

____

____

____

Experiment 8

Given: Two wires 10 or 15 inches long made of different metals; some lard or butter that is hard; candle or alcohol lamp.

Problem: To find out which one of the wires will let heat go through it the fastest. Suggestion: One of the wires could be of copper and one of iron.

How? ____

____

____

____

Experiment 9

Given: Some sand; balance.

Problem: To divide the sand into four equal piles.

How?

Experiment 10

Given: Two dry cell batteries; door bell; copper wire.

Problem: To connect both batteries to the bell so it will ring.

How?

Experiment 11

Given: Two strong sticks 12 or so inches long; a string about three inches long tied into a loop. The string should be so strong you cannot break it with your bare hands.

Problem: To break the loop of string.

How?.

PRECEDING PAGE MISSING

APPENDIX B

Special Adjustments for Lowest Ability Students

APPENDIX B

Special Adjustments for Lowest Ability Students

As is indicated in the body of the report, the reading level of our materials, as judged by the Dale-Chall Formula was generally too high and, in addition, was highly variable between units. It ranged from 4.0 to 5.0, but since only the more able groups went beyond the first three units during the one semester of work, it was in these units that we tried to reduce it in re-writing. In subsequent units it was up at sixth and seventh grade reading level, but the average to above students who had progressed rapidly and reached this far could handle this. But the problem was acute for the students of one group.

One of the seven elementary schools in Charlottesville was located in a somewhat deprived area; i.e., an area economically and culturally depressed in comparison with the other schools. It contained five of the nineteen classes which participated in Phase III (the evaluation of three major methods of teaching science during the spring of 1963). In addition, because these students were grouped by general academic ability and achievement, the lowest two classes of students were virtually non-readers. This meant that special help had to be extended to them -- that they could not participate in the experiment. They were given special help and these data were excluded from the study.

This was also the situation in that same school the following fall during Phase IV (the study of the three response factors in programed learning). At this time there were four classes instead of five. The data from the lowest two achieving classes were excluded from the analysis.

Throughout the rest of the system almost all students could participate. Those few who could not were given special help, but since they were randomly distributed throughout the testing sample, the results of the experiment were not affected.

The problem did not exist at all for the critical thinking phase because only the average to above students were included in the original design.

APPENDIX C

Initial and Final Means for
Eight Versions

APPENDIX C

Initial and Final Means for Eight Versions

	T1	T2	T3	T4	T5	T6	T7	T8
A	100.67	104.20	104.52	100.91	100.57	103.53	101.61	102.16
B	31.11	32.05	34.02	33.61	30.60	32.84	29.79	31.75
C	33.84	34.24	35.05	33.57	33.22	35.02	34.10	33.67
D	127.70	125.16	128.81	127.42	124.14	126.66	127.10	125.90
E	135.73	133.95	134.19	137.50	135.48	138.02	135.41	136.40
F	48.23	45.02	45.24	47.50	48.87	50.36	51.88	48.30
G	16.98	16.65	15.62	16.41	15.84	16.31	16.97	15.48
H	43.39	45.78	36.79	41.47	42.03	42.27	42.35	42.64

Where: A = I.Q.
B = Initial STEP
C = Final STEP
D = Initial Division True-False
E = Final Division True-False
F = Science Interest
G = Attitude
H = Interest Level

T1 = Constructed-Linear-All Answers
T2 = Constructed-Linear-1/3 Answers
T3 = Constructed-Branching-All Answers
T4 = Constructed-Branching-1/3 Answers
T5 = Multiple Choice-Linear-All Answers
T6 = Multiple Choice-Linear-1/3 Answers
T7 = Multiple Choice-Branching-All Answers
T8 = Multiple Choice-Branching-1/3 Answers

APPENDIX D

Percent Correct Responses Per Unit With No Responses Given

APPENDIX D

Percent Correct Responses Per Unit With No Responses Given

Class	Sound	Light	Heat	Science & Its Ways	Pl. & Anim.	Land, Water & Air
A	.67	.82	.92	.97	.94	.97
B	.61	.64	.78	.88	.77	.81
C	.53	.52	.51	.52	.52	.20
D	.75	.65	.71	.71	.75	.72
E	.80	.53	.65	.77	.77	.80
F	.86	.72	.82	.93	.93	.87
G	.87	.80	.88	.93	.93	.89
H	.72	.59	.78	.69	.65	---
I	.78	.67	.78	.83	.82	.80
J	.84	.83	.83	.61	.79	.88
K	.70	.52	.75	.85	.66	.81
L	.88	.58	.90	.93	.95	.91
M	.56	.69	.81	---	---	---
N	.55	.55	.54	.82	.63	---
O	.64	.53	.71	.79	.62	.72
P	.75	.71	.88	.91	.87	.95
Total	11.51	10.35	12.25	12.14	11.60	10.33
Mean	.72	.65	.77	.81	.77	.79

APPENDIX E

Branches Per Unit and Chi-Square Values For Each Branch

APPENDIX E

Branches Per Unit and Chi-Square Values for Each Branch

Unit - Sound

Branch	x^2
1	2.5
2	9.0*
3	12.9**
4	13.3**
5	10.4**
6	19.2**
7	10.4**
8	13.0**
9	10.3**
10	8.6*
11	3.4
12	12.1**
13	19.2**
14	4.3
15	9.7**
16	3.4
17	10.2**
18	12.5**
19	6.7*

Unit - Light

Branch	x^2
1	12.7**
2	11.6**
3	5.8
4	8.4*
5	10.5**
6	4.0
7	9.3**
8	9.4**

Unit - Heat

Branch	x^2
1	8.1*
2	5.9*
3	5.9*
4	9.4**
5	9.5**
6	6.7*
7	3.6

Unit - Science & Its Ways

Branch	x^2
1	5.7
2	8.6*
3	4.5
4	6.8*
5	5.4
6	11.1**
7	13.3**
8	4.7
9	6.0*
10	7.9*

Unit - Plants and Animals

Branch	x^2
1	5.6
2	2.9
3	8.5*
4	5.9*
5	4.6
6	4.9
7	3.7
8	4.3
9	2.6
10	5.0

Unit - Land, Water, and Air

Branch	x^2
1	11.2**
2	12.8**
3	8.8*
4	5.4
5	6.8*
6	6.8*
7	9.0*

*Significant at .05 level
**Significant at .01 level

APPENDIX F

Sample of Sound Critical Thinking

PRECEDING PAGES MISSING 1-10

SAMPLE OF SOUND CRITICAL THINKING

STATEMENT #2

You are driving in an automobile with a friend. The gasoline gauge on the dashboard is broken and you have been driving for several hours. Your friend says:

"The car is nearly out of gas."

You ask: "How do you know?"

He says: "Because I know how much gas was in the car when we started. I know how far we have gone. I also know how much gas this engine uses."

Is this good enough evidence for his conclusion?

________ (yes or no)

What objections to this evidence can you think of? Write them down here.

1. ____________________
2. ____________________
3. ____________________

Hint: It was good evidence so you did not need to write anything down.

It is good enough evidence, because if your friend does know all the things he says he knows, he can figure out how much gas is left.

STATEMENT #3

A farmer has built a new barn and stacked sacks of corn in it. You and a friend walk by to see it.

Your friend says: "The rats will get in and eat the corn."

You ask: "How do you know?"

Your friend says: "Because rats like corn and they can easily chew through these sacks."

Is this a good enough reason for his conclusion that the corn will be eaten by rats? ________ (yes or no)

Can you think of any objections to this reason? If you can, write them down here.

1. ____
2. ____
3. ____

Your main objections should have been that your friend does not know that the rats can get into the barn. The farmer may have put sheets of tin inside the walls to keep the rats from coming in.

PRECEDING PAGES MISSING 13-25

You have been learning about one way people think. Sometimes people are sloppy in their thinking. A good scientist tries to think carefully. Here is a story of two fourth grade boys talking. One of them is a good thinker. The other one is a sloppy thinker.

Joe said: "I don't like school."

Sam said: "What do you mean when you say you don't like school."

Joe said: "I mean just what I said. I don't like school!"

Sam said: "Do you like recess?"

Joe said: "Of course. Everybody likes recess. That's a silly question."

Sam said: "But you said you don't like school. Recess is part of school so that means you don't like recess either."

Joe said: "I mean I don't like school except for recess and playing ball and things like that."

Sam said: "Do you like history?"

Joe said: "NO."

Sam said: "Do you like science?"

Joe said: "Some of the science I like, but some of it I don't like. I like some of the experiments."

Sam said: "But you said you don't like school. Doing science experiments is school so that means you don't like science experiments either."

Joe said: "Okay! Okay! I still don't like school, but I do like recess and I like to do some of the science experiments."

Sam said: "Then your first statement was a little wrong, wasn't it? What should you have said instead of 'I don't like school.'?"

Joe said: "I guess, come to think of it, that I should have said -- ______________________________

____

____"

Sam said: "That is right. You should have said you do like some things in school and you don't like other things. Sometimes we say things we don't really mean exactly. We mean part of what we say, but we don't mean the other part." Here is an example.

PROBLEM: John did seven different experiments. Six of the experiments did not work. Only one of the experiments worked. John was a little angry. This is what he said:

CONCLUSION: "None of these darn science experiments work."

Was John's conclusion right? ________ (yes or no or maybe). What should he have said to be correct? __________

____

____

Here is another problem. See if you can figure this one out.

PROBLEM: The weatherman predicted rain today, but look! The sun is shining.

CONCLUSION: The weatherman never gets anything right.

Is the conclusion right? ________ (yes or no or maybe)

What might have been said instead? ______________________

____

____

____

PRECEDING PAGE MISSING

APPENDIX G

True-False Test

Critical Thinking Test

TEST

Directions: Read the statements carefully. If you think a statement is true, cross out the letter T on the answer sheet beside the number of that statement. If you think it is false, cross out the F beside the number of that statement.

Here is a chart.

	Tuesday	Thursday	Saturday
Morning	rain	fair	fair
Evening	rain	cloudy	rain

1. It rained all Tuesday, Thursday, and Saturday mornings.
2. It rained Saturday morning.
3. Tuesday evening it was clear.
4. It rained Saturday evening.
5. Thursday morning it was fair.
6. Thursday morning it was cloudy.
7. Green plants need carbon dioxide, sun light and water to make food.
8. Scientists still have many things to learn about the world.
9. Electricity is used to light our homes.
10. Living things need water to live.
11. One way to find out about something is to do an experiment.
12. Temperature is measured in degrees.
13. A magnifying glass will set a paper on fire because it brings many of the sun's rays together into a spot.

14. Sound travels in waves.

15. It is wise to keep our water and air clean for our good health.

16. Research is important in science.

17. A candle is a man-made light.

18. Observation is important in science.

19. Comparing two things is one way to study them.

20. Temperature is found by using a thermometer.

21. Sunlight is a kind of natural light.

22. Sounds are caused by vibrating objects.

23. A thermometer is used to measure temperature.

24. Science has helped change the world.

25. Electricity can produce heat in stoves and furnaces.

26. Seeds are carried to new places by wind.

27. Animals and plants have been changing over the years.

28. Science can help people make predictions.

29. We get heat from the moon.

30. All cells look alike.

31. Scientists search for evidence before they make conclusions.

32. A prism can separate sunlight into colors.

33. Roots take in water.

34. Many people put bread in boxes to keep the spores away.

35. Good written records are important in science.

36. Reflected sounds are called echoes.

37. Fish are able to move in the water by the use of their fins and tails.

38. We have day and night because the earth rotates.

39. There is soil in the air in the form of dust particles.

40. The largest animal on the earth today is the whale.

41. Light from the moon is reflected light.

42. There are many communities of living things in the oceans.

43. When water goes off into the air, we say it has evaporated.

44. When a piece of iron is heated, the molecules move faster.

45. The sun and the earth are about the same size.

46. The atmosphere is the ocean of air around the earth.

47. Grasshoppers are insects.

48. Heat can be produced by friction.

49. The two types of lenses are concave and convex.

50. The highness or lowness of sound is called pitch.

51. Sunlight is made up of seven basic different colors.

52. Sound cannot travel through the solid ground.

53. In order for electrons to flow, there must be a complete circuit.

54. Natural magnets can be found in the earth.

55. The earth turns on its axis.

56. Plants store food in roots, seeds, stems or leaves.

57. Some medicines are made from plants.

58. Water habitats would include, the oceans, streams and lakes.

59. A group of living things in its habitat is called a community.

60. Deserts have plenty of water.

61. Concrete sidewalks must have joints to take care of expansion and contraction due to temperature changes.

62. Friction is produced by one object moving back and forth over another.

63. Light bulbs are filled with air.

64. A magnet has invisible "lines of force" which go outside the magnet.

65. Chlorophyll is the green color in plants.

66. Glacier caps are very cold.

67. There is a complete circuit from the electric power plant to your house.

68. All birds can fly.

69. The study of the heavenly bodies is the science of astronomy.

70. Fossils tell us about plants and animals that lived long ago.

71. Chlorophyll is found in cells of green plants.

72. Man is made of cells.

73. The Indian learned to start fires by the use of friction.

74. Summer has the same amount of sunlight as winter.

75. It takes the earth about one year to revolve around the sun.

76. Water is made up of two chemicals called oxygen and hydrogen.

77. Plants which live in the desert are adapted to hold water.

78. The equator is the line that divides the northern part of the earth and the southern part.

79. Erosion is the washing or carrying away of the soil by wind and water.

80. The liquid part of the earth is called land.

81. Green plants make food in their leaves.

...

82. Animals can make their own food.

83. Chlorophyll is black.

84. You can see the molecules in water.

85. A mirror will reflect light in the form of an image.

86. Electricity can be stored as in a flashlight battery.

87. An electric switch is a device for making a circuit complete and incomplete.

88. It takes the earth 24 hours to make a complete turn on its axis.

89. Water is found in the air in the form of water vapor.

90. Leaves and twigs lying on the top of the ground help to stop erosion.

91. Stars are sources of natural light.

92. A vibrating object causes the air around it to vibrate.

93. Pressure is how much weight there is resting on a certain spot.

94. A word that means the same thing as measure is meter.

95. There is air in water and soil.

96. Vibrations can be heard but cannot be seen.

97. The United States is located in the northern half of the earth.

98. Molds need moisture.

99. Sound cannot travel through solid materials like iron or wood.

100. The light rays can be bent.

101. Slanting rays shine on the northern part of the earth when the earth tilts away from the sun.

102. All living things can be arranged in orderly groups.

103. A prism bends light.

104. Sound travels faster than light.

105. When electrons flow through a coil of wire, magnetism is produced.

106. A magnet can be made with electricity.

107. All animals are helpful to plants.

108. The amoeba can change the shape of its body.

109. All living things are either plants or animals.

110. A wet spot of water on the blackboard will dry faster on a wet day than on a dry day.

111. The roots of plants help to cause erosion.

112. Pitch is caused by speed of vibration.

113. When an object has extra electrons, we say it is negatively charged.

114. Magnets can be made from wood.

115. Lettuce has chlorophyll in it.

116. Molecules move faster in warm materials than they do in cold materials.

117. Water can bend light waves.

118. If air has the same temperature everywhere on the earth, the wind would blow.

119. There is a place for guessing in science.

120. A lens refracts or bends light.

121. A magnet has two poles.

122. Only green plants can make their own food.

123. Seaweed is made of algae.

124. Microscopes are used to look at stars.

125. Expansion is caused when molecules begin to move faster.

126. Direct rays of the sun are cooler than slanting rays.

127. A word that means the same as heat is thermos.

128. Pond scum is made of algae.

129. Sound causes molecules in a piece of iron to vibrate.

130. Man depends upon plants to help stop erosion.

131. Animals depend upon green plants for food.

132. Cells have to be seen through a telescope.

133. There is only one kind of electricity and it is called current electricity.

134. Air takes up space and has weight.

135. A habitat is the place where a plant or an animal lives.

136. The normal temperature of most people is about 78.6 degrees.

137. All living cells have chemicals in them.

138. The pressure down deep in the water is greater than the pressure at the top of the water.

139. When cold air comes into a room it rises towards the ceiling.

140. Bread mold has chlorophyll.

141. The color at the top of a rainbow is blue.

142. The part of the air that you breathe is called nitrogen.

143. Warm air rises and cold air will sink down.

144. When air gets cold it takes up more space than before.

145. When air is heated it expands or takes up more space.

146. There is more movement of molecules in solids than there is in liquids.

147. A campfire is man-made light.

148. Air has weight.

149. Water is a chemical.

150. When direct rays shine on the southern part of the earth, the southern part has less daylight than the northern part.

151. In a desert habitat there are plants and animals living near each other in communities.

152. From June to December the earth revolves all the way around the sun.

153. Electricity can be made with a magnet.

154. During June direct rays of the sun shine on the United States.

155. Penicillin is made from mold.

156. Science does not know the cure for cancer, but is searching for it.

157. Every object is made up of molecules.

158. You can see the smoke from a gun barrel before you can hear the sound of the gun.

159. When water is heated it will expand.

160. Science knows the cure for colds.

161. The cooling of a liquid causes it to shrink or contract.

162. A fuse is a safety device and helps to keep your house from catching on fire.

163. If the arrow on a windvane is pointing toward the south, the wind is coming from the south.

164. Spores float in the air.

165. If the point of a windsock is pointing toward the east, the wind is blowing from the east.

166. An amoeba is a one-cell animal.

167. Mushrooms and bread mold can make their own food.

168. All insects have six legs.

169. Desert habitats are hot or cold and dry.

170. A mold is a plant.

171. Living things breathe in oxygen to live.

172. A piece of iron gets larger when heated.

173. Sound travels faster through wood than through air.

174. Most thermometers are filled with red colored water.

175. When the earth makes a complete turn on its axis, it also makes a revolution about the sun.

176. The pressure of air is less on top of a mountain than it is down near the ground.

177. Birds help man by eating harmful insects.

178. Warm air pushes cold air out of the way.

179. Birds are adapted for flying by the use of wings, feathers, and light beaks.

180. Light travels faster than sound.

181. A rubber arrow sticks to the wall when the air pressure outside the rubber tip is greater than the air pressure inside the rubber tip.

182. Fish are adapted to live in the water by the use of their gills and scales.

Name ____________________ Teacher ____________________
Date ____________________ School ____________________

Directions:

Read the sentences carefully. Put a circle around yes if you think they show good thinking. Put a circle around no if you think they show bad thinking.

Example:

[✓ yes] no Bob is either running or walking. But Bob is walking. So, Bob is not running.

The yes has a circle around it because these sentences show good thinking.

1. yes no John is either in the fourth grade or in the third grade. John is in the third grade. So, John is in the fourth grade.

2. yes no The color of the book is either yellow or green. The color of the book is not green. So, it is yellow.

3. yes no The door is either closed or open. The door is open. So, the door is not closed.

4. yes no Susie is either reading or writing. But Susie is writing. So, Susie is reading.

5. yes no Jeff is either Tom's brother or Rose's brother. But Jeff is Rose's brother. So, Jeff is not Tom's brother.

6. yes no Jane is either sewing or talking or doing both at the same time. Jane is sewing. So, Jane is not talking.

7. yes no If all children go to school, then some children go to school.

8. yes no If some cars are small, then all cars are small.

9. yes no If some pupils have bicycles, then all pupils have bicycles.

10. yes no If all books are not green, then some books are green.

11. yes no If it rains on one Monday, then it will rain every Monday.

12. yes no If all policemen drive cars, then some policemen do not drive cars.

Directions:

Sometimes when we have two sentences, we can add a third sentence that shows good thinking.

Suppose we have these two sentences:

Example: All boys go to school.
Jack is a boy.

We can add another sentence.

[✓ yes] no All boys go to school.
Jack is a boy.
Therefore, Jack goes to school.

These three sentences show good thinking, so we put a circle around yes.

Here is an example of bad thinking. The circle is around no.

yes [✓ no] All dogs have four legs.
Tip is a cat.
Therefore, Tip has four legs.

Read the sentences below. If you think they show good thinking, put a circle around yes. If you think they do not show good thinking, put a circle around no.

13. yes no All dogs drink water.
Spotty is a dog.
Therefore, Spotty drinks water.

14. yes no All men wear pants.
All workers are men.
Therefore, all workers wear pants.

15. yes no All dogs drink water.
Spotty drinks water.
Therefore, Spotty is a dog.

16. yes no All girls are happy.

Jane is a girl.

Therefore, Jane is happy.

17. yes no All airplanes can fly.

All birds can fly.

Therefore, some birds are airplanes.

18. yes no All cows are animals.

All animals are living things.

Therefore, some cows are not living things.

19. yes no No birds are flowers.

All crows are flowers.

Therefore, no crows are birds.

20. yes no All workers are men.

Some farmers are men.

Therefore, some farmers are workers.

21. yes no All good children come early to school.

Some children ride their bicycles to school.

Therefore, all good children ride their bicycles to school.

22. yes no All fourth grade students are cute.

Bill and Susie are in the fourth grade.

Therefore, Bill and Susie are cute.

Directions:

You will find a few sentences below which tell you about a story. After the story, you will see some sentences. Next to each sentence you will see the words, yes and no.

Put a circle around the word yes if you agree with the sentence. Put a circle around the word no if you do not agree with the sentence.

A. Bill planted beans and corn seeds in a box. After some days these seeds came through the earth, because Bill

23. yes no a. kept the box in his room
24. yes no b. watered them every day
25. yes no c. covered them with a piece of paper
26. yes no d. kept the box in the shade of a tree
27. yes no e. put the box in the sunshine
28. yes no f. put the box in a cold place
29. yes no g. put the box in a warm place
30. yes no h. used good soil

B. Mary planted bean and corn seeds in the same box. After a few days only the corn seeds came through the earth. The bean seeds did not come through the earth. Some reasons why the bean seeds did not come through the earth are:

31. yes no a. covered seeds with too much dirt
32. yes no b. put the box in a warm place
33. yes no c. the sun was too bright
34. yes no d. may have planted bean seeds which were too old or too dry to grow
35. yes no e. put the box in a different place to get sunshine each day
36. yes no f. doesn't like the taste of beans

C. Suppose all the women who live in Huntsville wear red dresses and green hats. All the men wear green hats and red shirts. Everyone is called a child until he is 21 years old. Everyone who is 21 years or older is called a man or a woman. All the children who are not yet 21 years old wear red hats and red clothes.
Circle yes or no next to each sentence which talks about the people in Huntsville.

37. yes no a. All the people wear something red.

38. yes no b. None of the people wear something green.

39. yes no c. Everyone who wears red also wears a hat.

40. yes no d. All the people who wear red also wear something green.

41. yes no e. Some people wear something red.

42. yes no f. Everyone who wears a green hat is 21 years old or older.

43. yes no g. If you see someone wearing a red hat, this means that person is not yet 21 years old.

44. yes no h. All the men wear red shirts.

D. Four members of a family went on a picnic. Everyone ate ham sandwiches and potato salad. Betty and Bill picked apples from a tree and ate them. Father and Mother picked strawberries and ate them.

Now suppose something the family ate for lunch made them sick. Read the sentences below. Circle yes, no, or maybe.

45. maybe yes no a. Suppose Betty and Mother were sick. Betty said that the apples caused them to be sick. Do you think Betty is right?

46. maybe yes no b. Suppose Betty and Bill were the two people who got sick. Bill thinks that the apples caused them to be sick. Do you think Bill could be right?

47. maybe yes no c. Suppose all four members of the family got sick. Do you think the ham sandwiches caused their sickness?

48. maybe yes no d. Suppose Father and Mother got sick. Do you think the strawberries caused them to be sick?

E. Farmer Jones and Farmer Brown live next to each other. Their fields have the same kind of soil. During the summer, Farmer Jones and Farmer Brown grew corn. The corn needs 14 inches of rain to grow. It needs sunshine to grow, and 6 pounds of lime fertilizer for each field. The fields are near each other.
Farmer Jones' corn grew much better than Farmer Brown's corn. What do you think are some good reasons why Farmer Jones' corn grew much better than Farmer Brown's corn?

49. yes no a. Farmer Brown did not use fertilizer.

50. yes no b. Farmer Jones' field got more rain than Farmer Brown's field.

51. yes no c. Farmer Jones' field got more sunshine than Farmer Brown's field.

52. yes no d. Farmer Jones used better seed than Farmer Brown.

Directions: Sometimes we do not have what we need. We have to use something else.

Suppose you do not have a hammer.
If you want to hammer a nail, you would want to use
(a) a pillow (b) a piece of paper [✓ (c) a brick]

You would circle the word, brick. The pillow and piece of paper would not help you hammer the nail.

Circle the word that tells which object, or part of the object, you will use to help you.

53. If you want to make a needle you will use

(a) a fish (b) a book (c) a telephone

54. If you want to cut a piece of cake you will use

(a) an arm chair (b) a banjo (c) a pillow

55. If you want to sweep a floor you will use

(a) a chair (b) a pencil (c) a tree

56. If you want to make eyes for a doll you will use

(a) a shirt (b) a stocking (c) a desk

57. If you want to light a match you will use

(a) a book (b) a sponge (c) a ball

58. If you want to make a cover for your book you will use

(a) a tin pan (b) a bag (c) a desk